University of Stirling

Faculty of Natural Sciences

Division of Computing Science and Mathematics

Innovative Signal Processing and Data Mining Techniques for Aquatic Animal Health

Alexander Frederick Bethune Carmichael

A thesis submitted in fulfilment
of the requirements for the degree of
PhD in Computing Science

September 2023

Thesis Committee

Newcastle University

Dr. Deepayan Bhowmik

University of Stirling

Prof. Gabriela Ochoa

Dr. Johanna L. Baily

Prof. Jimmy Turnbull

Scotland's Rural College (SRUC)

Dr. Annette S. Boerlage

Prof. George J. Gunn

Dr. Aaron Reeves

Dr. Andrew Brownlow

Abstract

Problem: Aquatic animal health data is often stored in unstructured formats like text and medical images, making large-scale analysis challenging due to the complexity of processing such data.

Objectives: In this thesis, we aim to develop text mining, signal processing, image processing, and machine learning techniques to analyse unstructured data effectively. These methods will enable the aggregation of information across large datasets of unstructured aquatic animal health data.

Methodology:

- For text analysis, we have designed an ontology-based framework for extracting and storing information from aquatic animal post-mortem reports, with a focus on gross pathology reports. While we initially applied this framework to marine mammal stranding reports, it can be adapted for various species and report types.
- For medical image analysis, we have created methods for identifying and analysing lesions in whole-slide images (WSIs) of Atlantic salmon gills. Our approach includes a novel feature extraction technique utilising the empirical wavelet transform, and we enhance context-awareness by employing a variational autoencoder to identify regions of interest within histology images.

Achievements: The research resulted in the development of an ontology-based framework for systematic text extraction and storage from marine mammal gross pathology reports. We showcased our framework's performance by using it to analyse bottlenose dolphin attacks on harbour porpoises. Additionally, we created innovative methods for lesion detection in Atlantic salmon gill whole-slide images, incorporating advanced techniques such as the empirical wavelet transform, deep learning, and a variational autoencoder for context-awareness. These achievements collectively advance the analysis of unstructured aquatic animal health data, enabling more comprehensive and efficient data processing. At the time of writing, the project is the only one to apply data-driven approaches to marine mammal post-mortem reports and gill WSIs.

Attestation

I understand the nature of plagiarism, and I am aware of the University's policy on this.

I certify that this dissertation reports original work by me during my PhD except for the following:

- The Littlewood-Paley Empirical Wavelet Transform code used in Chapter 6 was created by Basile Hurat (https://github.com/bhurat/EWT-Python).
- The stain normalisation code was implemented as part of the torchstain library (https://github.com/EIDOSLAB/torchstain) [1, 2].

Signature:  **Date:** 30/09/2023

List of Publications

This thesis has produced two peer-reviewed publications:

1. [3] Carmichael, A.F., Bhowmik, D., Baily, J., Brownlow, A., Gunn, G.J. and Reeves, A., 2020, September. **Ir-Man: An Information Retrieval Framework for Marine Animal Necropsy Analysis.** In Proceedings of the 11th ACM International Conference on Bioinformatics, Computational Biology and Health Informatics (pp. 1-9).

2. [4] Carmichael, A.F., Baily, J.L., Reeves, A., Ochoa, G., Boerlage, A.S., Gunn, G., Allshire, R. and Bhowmik, D., 2023, April. **Analysing hyperplasia in Atlantic salmon gills using empirical wavelets.** In Medical Imaging 2023: Digital and Computational Pathology (Vol. 12471, pp. 116-123). SPIE.

Acknowledgements

I want to express my deepest gratitude to all those who have supported me throughout my doctoral journey. This thesis would not have been possible without the unwavering encouragement, guidance, and assistance of numerous individuals and institutions.

First and foremost, I am profoundly indebted to all members of my supervision team. Thanks to Dr. Deepayan Bhowmik for encouraging me to pursue a research career. His mentorship has been invaluable throughout this endeavour. I want to thank Dr. Annette Boerlage, Dr. Johanna Baily, Dr. Aaron Reeves, Prof. Gabriela Ochoa, Prof. George Gunn, Dr. Andrew Brownlow, and Prof. Jimmy Turnbull for their support, guidance, and supervision, enabling this truly interdisciplinary project to thrive.

I am deeply grateful to the University of Stirling and Scotland's Rural College (SRUC) for providing me with the resources, research facilities, and financial support that enabled me to pursue my doctoral studies.

I extend my sincere gratitude to Dr. Johanna Baily and Rosie Allshire for scanning and provision of slides from Cooke Aquaculture. Their support greatly facilitated the initial stages of image processing in this project. My appreciation also goes to Chris Matthews and the dedicated team at Pharmaq Analytiq (formerly Fish Vet Group) for providing the final dataset, and I would like to thank Mowi Scotland, Bakkafrost and Loch Duart for sharing their data. Additionally, I would like to thank Dr. Andrew Brownlow, Mariel ten Doeschate, and the rest of the Scottish Marine Animal Strandings Scheme (SMASS) team for their invaluable assistance, particularly for providing the reports that were instrumental during the project's first year.

I want to thank Sam and Graham of the Computing Support Group at Stirling for the invaluable technical assistance that kept my research running smoothly. Thanks also to Grace and Gemma of the admin team for their administrative support and assistance throughout my doctoral journey.

To my family, your unwavering support, patience, and belief in my abilities have been my constant source of motivation.

I extend my heartfelt appreciation to the friends who have stuck by me throughout the challenging past few years. Their understanding, encouragement, and occasional distractions have provided the balance I needed during this demanding period.

I want to dedicate a special thank you to my fiancé, Becky. Her love, boundless patience, and steadfast support were the bedrock upon which this thesis was made possible. Finally, heartfelt thanks go to my feline companions, Gatsby and Pearl, who offered occasional and transactional support—particularly when their appetites weren't the priority.

Key Terms and Abbreviations

- *Pathology*- The study of disease.
- *Histology*- The microscopic study of the cells, tissues, and organs.
- *Histopathology*- The microscopic study of the disease in cells, tissues, and organs.
- *Lesion*- A region of damaged tissue.
- *Hyperplasia*- Increase in the amount of tissue due to cell proliferation.
- *Hypertrophy*- A lesion involving an increase in the size of individual cells.
- *Epithelium*- One of the basic types of tissue. It provides a thin layer of protection on the surface of many organs.
- *Mucous Cells*- Cells which produce mucus which protects tissue from abrasion.
- *Capillary*- The smallest type of blood vessel that connects arterioles with venules.

Abbreviation	Meaning
AI	artificial intelligence
AUC	area under curve
BDA	bottlenose dolphin attack
CAD	computer-assisted diagnosis
CDC	Centre of Disease Control
CGD	complex gill disease
CNN	convolutional neural network
DT	determiner
EMD	empirical mode decomposition
EMR	electronic medical record
EWT	empirical wavelet transform
FCNN	fully connected neural network
FFT	fast fourier transform
FN	false negative
FP	false positive
GLCM	gray-level co-occurrence matrix
GMM	Gaussian mixture model
GSA	grey seal attack
H&E	hematoylin and eosin
JJ	adjective
LBP	local binary patterns
LP	Littlewood-Paley
ML	machine learning
NAD	no abnormalities detected
NE	not examined
NLTK	Natural Language Toolkit
NN	neural network
NP	noun phrase
OCR	Optical Character Recognition
OD	optical density
PCB	polychlorinated biphenyls
PDF	probability density function
PM	post-mortem
PMI	pointwise mutual imformation
ROC	receiver operating characteristic
ROCO	Radiology Objects in Context
ROI	region-of-interest
SMASS	Scottish Marine Animal Strandings Scheme
SVM	support vector machine
TN	true negative
TP	true positive
UMLS	Unified Medical Language System
VAE	variational autoencoder
VB	verb
WSI	whole slide image

Contents

List of Figures

1 Introduction

1.1 General Context and Motivation

The medical domain has seen significant advancements in recent years with the rise of automated analysis methods, particularly machine learning (ML) [5]. ML has revolutionised how medical data is analysed and interpreted, improving accuracy and efficiency in diagnosis and treatment planning. This has paved the way for developing innovative applications and tools that have significantly improved patient outcomes and revolutionised the healthcare field [6].

Although artificial intelligence (AI) technologies have been widely implemented in the human medical field [5], there is an opportunity to leverage its potential in veterinary medicine. By using machine learning algorithms to analyse legacy data, including post-mortem reports and patient records, pathologists can gain new insights into animal health and behaviour. These new insights could lead to interventions or policy changes which could significantly impact animal health and welfare, and ultimately benefit the veterinary industry as a whole. [7]

Observing and monitoring the health and well-being of a population is essential for identifying factors that may have a significant impact on it. However, such tasks can be challenging, especially for aquatic animals. Unlike land animals, observing aquatic creatures in their natural habitat is often unfeasible, making it harder to keep tabs on their current state of being.

Thankfully, post-mortem (PM) analysis can provide a rare and valuable

insight into the condition of animals. This is the case for both wild and cultured animals. Monitoring their living state is a difficult task in both cases, however, PM analysis allows researchers to obtain valuable information that can help build a greater epidemiological understanding of the population. By aggregating the data produced from this analysis, researchers can gain deeper insights into the health and well-being of aquatic animals, even in the absence of live observation.

Two of the most prevalent methods for storing post-mortem data are through text, often used for reporting findings and medical images of samples. By incorporating unstructured modalities such as gross pathology reports and medical images into epidemiological analysis, a better understanding of the population can be obtained. PM analysis of marine mammals, such as cetacea and pinnipeds, has already been shown to give an epidemiological understanding of marine mammal populations. Williams *et al.* [8, 9], for example, have extensively monitored the levels of toxic polychlorinated biphenyls (PCBs) in harbour porpoises. Human pollution directly affects PCB levels due to the compound's use in some manufactured goods. Similarly, Nelms *et al.* [10] analysed the presence of microplastics found in stranded cetacea using PM examinations. These examples show that PM examinations can be used to observe the human impact on marine ecology.

Such analysis methods have also been used to improve the field of aquaculture, with PM analysis being used to study phenomena such as antimicrobial resistance in marine-farmed salmon [11] and the impacts of dietary and environmental changes [12].

Such insights have a significant impact providing several clear motivations. Some benefits are common to both the marine mammal and aquaculture domains:

- **Epidemiological Understanding**: The ability to aggregate unstructured data sources is crucial for effective disease profiling. By collecting and organising data from various sources, researchers can gain a more comprehensive understanding of the disease, its progression, and the factors contributing to its development. The quantification of pathological features can allow for analysing trends and patterns that would otherwise be difficult to identify. By examining these trends, researchers can improve their epidemiological understanding of the disease and

gain insights into potential treatments or preventative measures. This analysis can be a powerful tool to help identify risk factors and develop strategies to reduce disease prevalence, ultimately improving health outcomes. Thus, these methods can potentially impact disease prevention and management significantly.

- **Pathological Understanding**: Quantifying the level of pathology in unstructured modalities, particularly medical images, can offer numerous benefits for pathological understanding, including standardisation, objectivity, and pathological profiling. Such an approach can provide a consistent and objective approach to evaluating tissue samples, making results comparable and reproducible across different research studies. Additionally, quantification reduces the potential for subjective interpretations that can be influenced by personal biases.
- **Animal Welfare**: By analysing unstructured modalities data, stakeholders can develop targeted interventions to improve environmental conditions, such as reducing pollutant levels. These interventions can enhance animal welfare and overall quality of life. For marine mammals, this understanding will typically surround the impact of human activity on the animal's health and outcomes, allowing researchers to study phenomena such as entanglement and bycatch [13].

Some benefits are more specific to the field of aquaculture due to the field's commercial nature:

- **Sustainability**: Improvements to animal welfare and environmental conditions in aquaculture can positively impact the surrounding ecosystem. By ensuring that animals are kept in healthy and sustainable conditions, aquaculture operations can reduce the risk of disease outbreaks and the need to use antibiotics or other chemicals that can negatively impact the environment [14]. Providing animals with optimal conditions for growth and health can reduce the amount of waste produced, leading to more efficient utilisation of time and resources.
- **Food Security**: By analysing trends and patterns in the health of animal populations, stakeholders can identify potential risks and implement interventions to prevent disease outbreaks or other health-related events that could disrupt the supply chain. This is especially pertinent given the impact that climate change will have on food security across

the board [15], and the important role that aquaculture is going to play in ensuring food security, given that 178 million tonnes of aquatic animals was produced globally in 2020 and has been steadily increasing over the last 50 years [16].

1.2 Aim and Objectives

The main objective of this project is to develop data-driven methods to extract and aggregate information from unstructured data and develop novel methods that allow for statistical analysis across multiple data modalities. The primary aim is to develop tools that support understanding aquatic animals' health and condition by considering various data sources, including unstructured modalities such as medical images and free-text pathology reports, which may be used alongside structured fields such as date and location in future. Data produced by these methods can provide a comprehensive and accurate analysis of animal health, which could provide valuable insights for researchers and stakeholders in animal health domains.

This project has two primary objectives:

- To develop an information retrieval framework for aquatic animal gross pathology free text reports to demonstrate the value of automating the analysis of pre-existing necropsies.
- To create novel computer vision methods to extract lesion data from aquatic animal medical images.

1.3 Contributions

The main contributions of this work are:

- The analysis of literature relating to information retrieval techniques for biomedical texts.
- The analysis of literature relating to automated analysis of histology slide images, particularly those relating to fish tissue.
- The development of an ontology-driven information retrieval framework known as Ir-Man.

- A demonstration of Ir-Man's performance by applying the framework to harbour porpoise gross pathology reports.
- The development of an automated approach for quantifying a lesion known as hyperplasia in Atlantic salmon in gill Whole-Slide Images (WSIs).
- The development of a novel empirical wavelet-based method for feature extraction, which uses probability density functions to generate features.
- A proof-of-concept tile-based dataset produced using a sliding window approach for hyperplasia severity classification.
- A region-of-interest classification component that utilises variational autoencoder (VAE) based anomaly detection.
- Metrics for hyperplasia prevalence in gill histology images.
- Heatmap visualisations based on the metrics provided by our models.
- A comparison of the metrics generated by our approach to manually annotated scores generated by expert histopathologists.

1.4 Research Questions

To fulfil the aims and objectives of this project, several research questions have been identified:

- *How should we develop robust text mining techniques for extracting meaningful information from aquatic animal post-mortem texts?*
- *What methods already exist for the automated analysis of aquatic animal histology images?*
- *What image processing techniques should be developed for the automated analysis of aquatic animal histology images?*
- *How well do the models developed in this project perform compared to expert histopathologists?*

1.5 Thesis Organisation

Chapter 2 provides background information on the marine mammal strandings and aquaculture domains, with an initial focus on the current states of the domains and some context information on the domain terms used in

the project. The latter section of the chapter discusses how AI and other automated methods are currently used to analyse aquatic animal health.

Chapter 3 is a state-of-the-art analysis of automated information retrieval methods in text and imaging. This involves a comprehensive study of techniques such as text mining and its application in biomedical texts, particularly those in clinical settings and in the context of post-mortem analysis. The chapter then discusses how anomaly detection methods are currently applied in medical imaging, particularly those which incorporate variational autoencoders. Lastly, the state-of-the-art chapter discusses signal decomposition methods and their application in the medical imaging domain.

Chapter 4 describes the framework developed to extract relevant information from marine mammal gross pathology reports. We describe the language used in these reports before outlining how we use tailor-made ontologies as a guideline for extracting observations pertaining to anatomical entities. We then demonstrate the effectiveness of our approach on a use case involving bottlenose dolphin attacks on harbour porpoises.

Chapter 5 lists and describes the image datasets used in Chapters 6, 7, and 8.

Chapter 6 provides an overview of our strategy for hyperplasia analysis in Atlantic salmon gill whole-slide images. We formulate the problem as a classification task and outline our preprocessing steps, which include stain normalisation techniques. Next, we describe our innovative approach to feature engineering, which involves generating parametric features from subband images generated via Empirical Wavelet Transforms. Additionally, we cover the training and evaluation of multiple models, incorporating both this method and established deep learning techniques.

Chapter 7 improves upon the context-agnostic approach introduced in Chapter 6 by eliminating irrelevant regions that do not contribute to the hyperplasia analysis task. We achieve this by treating these irrelevant regions as anomalous tissue, thereby enabling the utilisation of anomaly detection techniques based on variational autoencoders.

Chapter 8 assesses the efficacy of various elements within our image processing pipeline by contrasting the scores it generates with those derived from expert histopathologists, which serve as representations of global whole-

slide image hyperplasia scores. We investigate potential aggregation metrics and create visualisations to facilitate a detailed discussion of our model's performance.

In **Chapter 9** begins with a discussion of the key findings from our research and their implications for industry, policy, and future research. We then address the limitations of our methods and analysis, providing recommendations for further research in the field.

Chapter 10 offers a summary of the thesis' work and formally presents the recommendations for future research in the domain, building upon the suggestions outlined in the preceding chapter.

2 Background

Post-mortem analysis has been a valuable tool in supporting our understanding of anatomy and pathology for centuries. The practice dates back to ancient times, with the earliest recorded autopsies performed by the Greek physician Galen of Pergamon (131-200 BC) [17]. Galen was a pioneer in anatomy and pathology research and made significant contributions to our knowledge of the human body. He was the first to establish a connection between the ailments described by patients and the diseased organs discovered during post-mortem observations. Galen's findings provided essential insights into the causes of disease and helped shape the development of medical knowledge. In the centuries that followed, post-mortem analysis became an increasingly common practice in medicine, providing valuable information on the structure and function of organs and the effects of disease on the body [18].

Post-mortem analysis of animals (also known as a necropsy) was likely first performed for practical purposes, such as determining which organs of an animal were safe for consumption. Later work by Galen and other physicians, such as Herophilus and Erasistratus, would lead to significant advances in comparative anatomy and helped establish the basic principles of modern veterinary medicine [18]. Performing necropsies and aggregating the findings has also provided valuable information on the prevalence and spread of diseases in animal populations, allowing for improved diagnosis so that stakeholders can better mitigate disease and protect both animal and human health.

In this chapter, we will first describe the domains of marine mammal

strandings and gill health with a focus on applications in aquaculture. The selection of these two groups is based on their distinctive features, including the challenges they present regarding epidemiological surveillance and the availability of underutilised post-mortem data. By exploring these subdomains, this study seeks to advance our knowledge of aquatic animal health and contribute to developing novel techniques for both images and texts as tools to support diagnosis. We will then explore the current challenges facing aquatic animal health and the traditional methods for gaining epidemiological understanding. These conventional methods include studying disease prevalence and mortality rates in wild populations and underlying causes of premature death in aquaculture. We will also address recent technological advances and the increasing availability of high-quality data that have led to the development of state-of-the-art methods for studying aquatic animal health. Lastly, we will identify new potential areas where applying state-of-the-art techniques (including data mining and machine learning) could significantly impact.

2.1 An Overview of Aquatic Animal Health

The primary focus of this research is to explore the potential of automation in enhancing pathological and histological analysis in aquatic animal health. The study will utilise advanced computational tools and cutting-edge technologies to investigate how automation can revolutionise these critical domains. This research's findings will help enhance our ability to monitor and diagnose aquatic animal diseases in real time, leading to better disease control and prevention.

It is essential to note that the techniques developed and described in this thesis are not limited to marine mammals and farmed fish but can be applied in other areas of medicine and fisheries contexts. Thus, this study has the potential to contribute to the broader field of animal health by introducing innovative approaches to pathological and histological analysis.

2.1.1 Marine Mammal Strandings

The collection and organisation of data from marine animal stranding events in Scotland are carried out by the Scottish Marine Animal Stranding Scheme (SMASS). A stranding event is when one or more marine animals beach themselves on a shore or in shallow waters. These events can occur for various reasons, such as illness, injury, disorientation, or confusion caused by noise pollution or changes in the ocean's temperature, among other factors. In many cases, stranding events are fatal for the animal(s) involved. One of SMASS's primary focuses is the post-mortem (PM) analysis of stranded marine mammals. Although SMASS documents many species, including basking sharks and marine turtles, their work predominantly deals with cetaceans (such as harbour porpoises, bottlenose dolphins, and whales) and pinnipeds (seals).

When a deceased, stranded animal is subjected to PM analysis, a detailed report is produced which provides an in-depth description of the analysis carried out and the animal's condition. Typically, these reports contain conclusions about the animal's cause of death. PM reports are semi-structured, with both structured and unstructured free text fields. In this work, the terms unstructured text, and free text are used synonymously. The structured fields are described in Tables 2.1 and 2.2. Figure 2.1 shows the first page of an example PM report.

Current Data Science Applications in Marine Animal Health

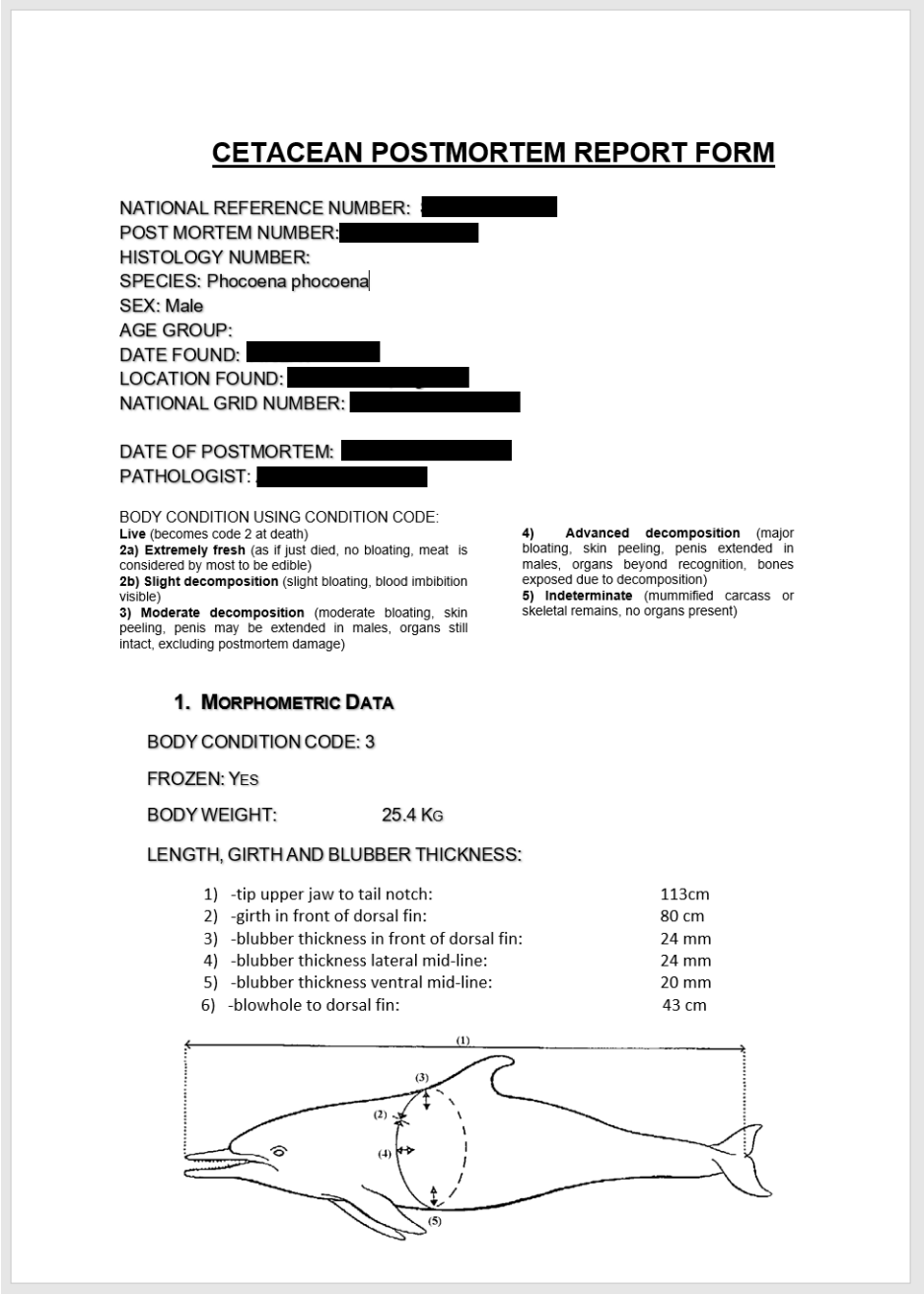
As with many other fields, analysis of marine mammal health is seeing a significant rise in the development of data science-based methods due to the advancements in, and increasing availability of deep learning and other computer vision techniques.

Aditya Jyoti Paul et al. conducted an in-depth analysis of recent advancements in the detection and identification of several sea turtle species using cutting-edge computer vision techniques. They identified a number of CNN methods employing a variety of imaging types, including those obtained by aerial drones and autonomous underwater vehicles [19].

In 2019, Fretwell et al. performed a case study where they used satellite

Field	Description
NATIONAL REFERENCE NUMBER	A national referencing system is used across the UK. Each stranding PM report has a unique reference number.
POST MORTEM NUMBER	A unique reference number used within the SMASS system.
HISTOLOGY NUMBER	A unique reference number identifying a histopathology report (where one exists).
SPECIES	Taxonomic to which the animal belongs. May be in English or Latin.
SEX	The sex of the animal. May also identify if the animal is pregnant.
AGE GROUP	A general description of the animals age (e.g. 'juvenile', 'adult')
DATE FOUND	The date that the animal is reported to SMASS
LOCATION FOUND	The name of the approximate location that the animal was found.
NATIONAL GRID NUMBER	A national grid reference coordinate identifying the animal's exact location.
DATE OF POST-MORTEM	The date that PM analysis was performed.
PATHOLOGIST	The investigator(s) responsible for analysis and production of the PM report.
MORPHOMETRIC MEASUREMENTS	Multiple fields which describe different morphometric features of the animal, including length, girth and blubber thickness. Also includes the animal's weight and whether the animal was frozen for PM analysis.
BODY CONDITION CODE	A code which describes the overall carcass condition. This is impacted by the level of autolysis and the presence of scavenger damage.
CAUSE OF DEATH	The likely ultimate or proximal cause(s) of death.

Table 2.1: Structured fields used in SMASS PM reports.



CETACEAN POSTMORTEM REPORT FORM

NATIONAL REFERENCE NUMBER: [redacted]
POST MORTEM NUMBER: [redacted]
HISTOLOGY NUMBER:
SPECIES: Phocoena phocoena
SEX: Male
AGE GROUP:
DATE FOUND: [redacted]
LOCATION FOUND: [redacted]
NATIONAL GRID NUMBER: [redacted]

DATE OF POSTMORTEM: [redacted]
PATHOLOGIST: [redacted]

BODY CONDITION USING CONDITION CODE:
Live (becomes code 2 at death)
2a) Extremely fresh (as if just died, no bloating, meat is considered by most to be edible)
2b) Slight decomposition (slight bloating, blood imbibition visible)
3) Moderate decomposition (moderate bloating, skin peeling, penis may be extended in males, organs still intact, excluding postmortem damage)
4) Advanced decomposition (major bloating, skin peeling, penis extended in males, organs beyond recognition, bones exposed due to decomposition)
5) Indeterminate (mummified carcass or skeletal remains, no organs present)

1. MORPHOMETRIC DATA

BODY CONDITION CODE: 3

FROZEN: YES

BODY WEIGHT: 25.4 KG

LENGTH, GIRTH AND BLUBBER THICKNESS:

1)	-tip upper jaw to tail notch:	113cm
2)	-girth in front of dorsal fin:	80 cm
3)	-blubber thickness in front of dorsal fin:	24 mm
4)	-blubber thickness lateral mid-line:	24 mm
5)	-blubber thickness ventral mid-line:	20 mm
6)	-blowhole to dorsal fin:	43 cm

Figure 2.1: The first page of an example SMASS PM report. General structured information is shown, while unstructured and semi-structured text is shown on later pages.

Field	Description
GROSS PATHOLOGY REPORT	A report describing the characteristics of the body as a whole, as well as different anatomical parts of the carcass.
BACTERIOLOGY REPORT	A report of the results of a bacteriological examination of any samples taken.
HISTOPATHOLOGY REPORT	A report of the histopathological examination of tissue samples taken from the carcass. Each tissue region tends to be described in isolation.
OBSERVATIONS/COMMENTS	A general summary of the findings of the PM analysis.

Table 2.2: Unstructured fields used in SMASS PM reports.

imagery to detect stranded whales in Chile. They evaluated both manual and automated methods and recommended that future work in the field should involve machine learning methods.

Despite the recommendations of Fretwell et al., a review in 2022 found that there are still significant gaps in the field with most automated detection methods focussing on baleen whales and pinnipeds [20].

With regard to text mining of marine mammal data, very little work currently exists. A 2021 study by Coram et al. analysed facebook posts using simple keyword search-based methods to analyse the impact of litter on cetaceans in Southeast Asia [21]. This work was later criticised by Peter et al. for only using a few keywords on one social media platform, and for focusing on English text, meaning that the conclusions reached were flawed [22]. This highlights the need for more advanced data analysis methods and careful text selection from which we gain our conclusions.

2.1.2 Aquaculture

Aquaculture is the practice of growing aquatic organisms under controlled conditions, including fish, shellfish, algae, and plants. Aquaculture is a significant and growing sector in Scotland that produces food, jobs, and exports. The main aquaculture product in Scotland is Atlantic salmon, the with Scotland being the third largest producer in the world and the country's top food export [23]. Other finfish species in Scotland, such as rainbow trout, halibut,

and cleaner fish, are farmed in the sea, whereas shellfish, such as mussels, oysters, and scallops, are cultivated along the coastline. There is also potential for seaweed farming, which can offer various benefits such as biofuels, pharmaceuticals, and fertilisers [24]. This section describes the challenges currently faced by the domain and provides context for some of the domain terminologies and concepts used throughout this thesis. Lastly, a brief description of the data sources is provided.

Current Challenges for Aquaculture

As with many industries associated with food security, aquaculture in Scotland and globally, faces a wide variety of challenges:

- Environmental impacts: Aquaculture can impact the freshwater and marine ecosystems' health, biodiversity, and water quality. [25] For instance, finfish farms' waste, chemicals, and parasites can harm wild fish and other organisms. Shellfish farms can also change the physical and biological characteristics of the seabed. Stakeholders must adhere to strict regulations and guidelines to minimise and manage their environmental effects.
- Climate change: Due to factors like rising ocean temperatures, acidification of the ocean, extreme weather, and disease outbreaks, aquaculture is susceptible to the effects of climate change [26]. These increase the dangers of escapes and interactions with wild populations while also impacting aquaculture species' growth, survival, and health. Aquaculture also adds to greenhouse gas emissions through its use of energy, feed, and transportation. Aquaculture must adapt to climate change and reduce carbon emissions to ensure its sustainability. [27]
- Social and economic factors: Providing a skilled and sufficient workforce, adequate infrastructure and services in rural areas, upholding high standards for animal welfare and product quality, satisfying consumer demand and preferences, competing with other sectors for space and resources, and adjusting to market fluctuations and uncertainties are just a few of the challenges faced by aquaculture [28].

Histology Overview

As with many other image processing and computer vision tasks, automated histology analysis usually consists of classification, detection and segmentation tasks. Classification tasks typically involve Computer Assisted Diagnosis (CAD) of pathologies such as tumours. Detection tasks in the context of histological analysis are generally concerned with the counting and isolation of certain types of cells or nuclei. Lastly, segmentation tasks are concerned with creating superpixels of cell types or tissues [29].

Creating histology images is essential to medical research and diagnostics, as it allows scientists and pathologists to examine the microscopic structure of tissues and cells. These images provide valuable insights into an animal's health or disease state and play a crucial role in understanding the underlying mechanisms of various medical conditions.

1. **Tissue Sample Collection**: The first step in creating histology images involves obtaining a tissue sample from a patient. Depending on the specific diagnostic or research objectives, this can be done through various methods, such as biopsy or surgical excision. Handling the tissue sample carefully is essential to preserve its integrity and minimise artefacts.
2. **Tissue Fixation**: Immediately after collection, the tissue sample is fixed to preserve its cellular and structural components. Formalin, a solution of formaldehyde, is commonly used for fixation. Fixation helps prevent decay and autolysis, maintaining the tissue's cellular structure for analysis.
3. **Tissue Processing**: Once fixed, the tissue sample goes through processing steps, including dehydration, clearing, and embedding in paraffin wax.

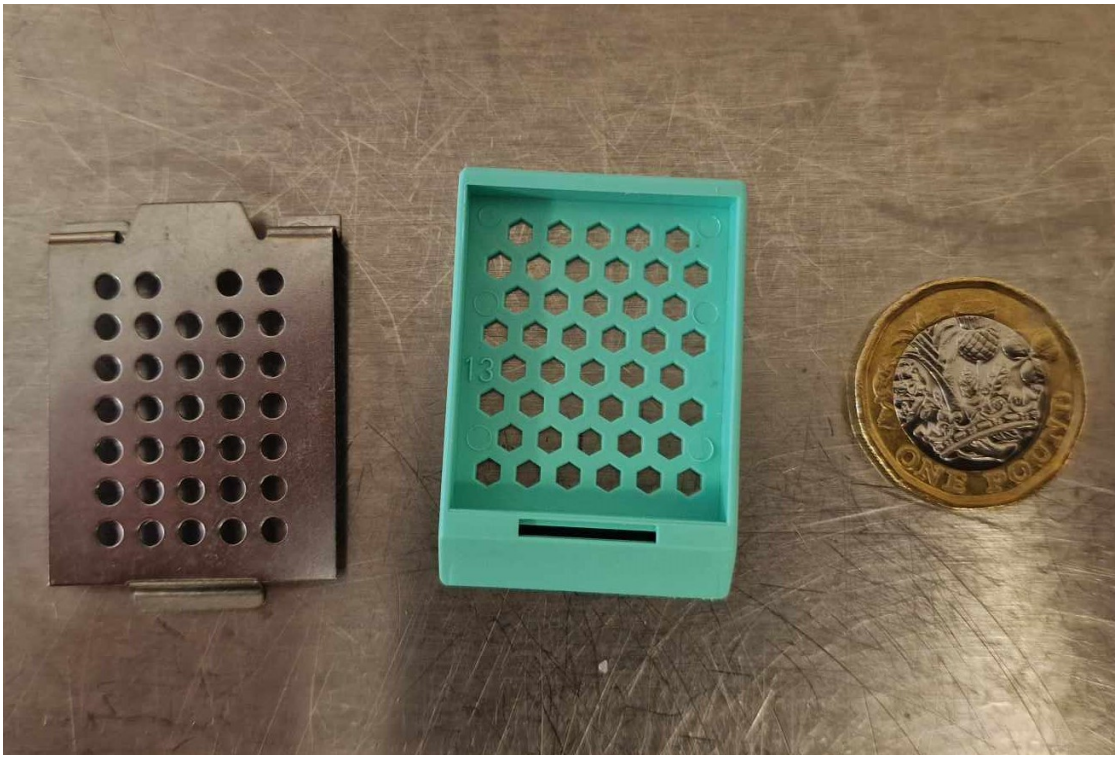
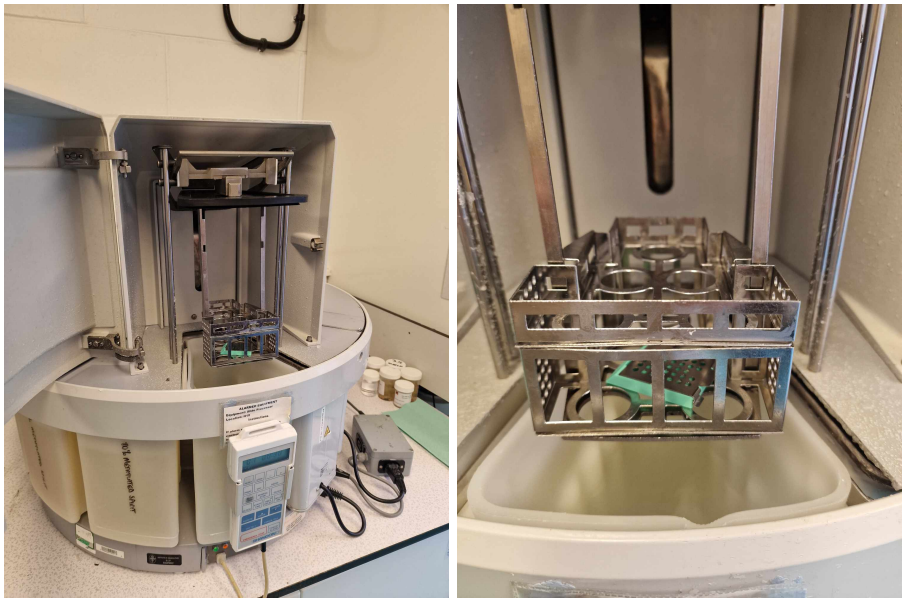


Figure 2.2: A cassette, which is used to hold the sample through tissue processing next to a £1 coin for scale.

The tissue is mounted in a “cassette", shown in Figure 3.1, before being placed in a tissue processor, such as the machine shown in Figure 2.3. These steps remove water from the tissue and make it suitable for sectioning. Dehydration is typically done using a series of alcohol solutions of increasing concentrations.



(a) Tissue Processor (b) Cassette.

Figure 2.3: Figure (a) shows a SHANDON Citadel 2000 tissue processor, and (b) shows a cassette held in it. This process involves dehydration, clearing, and embedding in paraffin wax.

4. **Tissue Sectioning**: The processed tissue is then sliced into thin sections, typically 5 micrometers thick, using a microtome (shown in Figure 2.4 alongside other crucial equipment such as a water bath, hot plate, and a cold plate). These sections are so thin that they are transparent to visible light. The quality of sectioning is crucial to obtaining clear and informative histology images.
5. **Mounting**: The thin tissue sections are mounted onto glass slides. A small amount of adhesive, such as a gelatin-based solution, affixes the tissue to the slide.

Figure 2.4: A workstation containing a water bath, microtome, hot plate, and cold plate. Many of the slides used in this work have been produced in a similar setting.

6. **Staining**: Histological staining is a critical step that contrasts the tissue sections, making cellular and structural features more visible under a microscope. An example of the final slide can be seen in Figure 2.5. In Hematoxylin and Eosin (H&E) staining which is the focus of this work, nuclei and ribosomes are stained in shades of blue and purple, while cytoplasm and collagen tend to be stained in shades of pink and red.

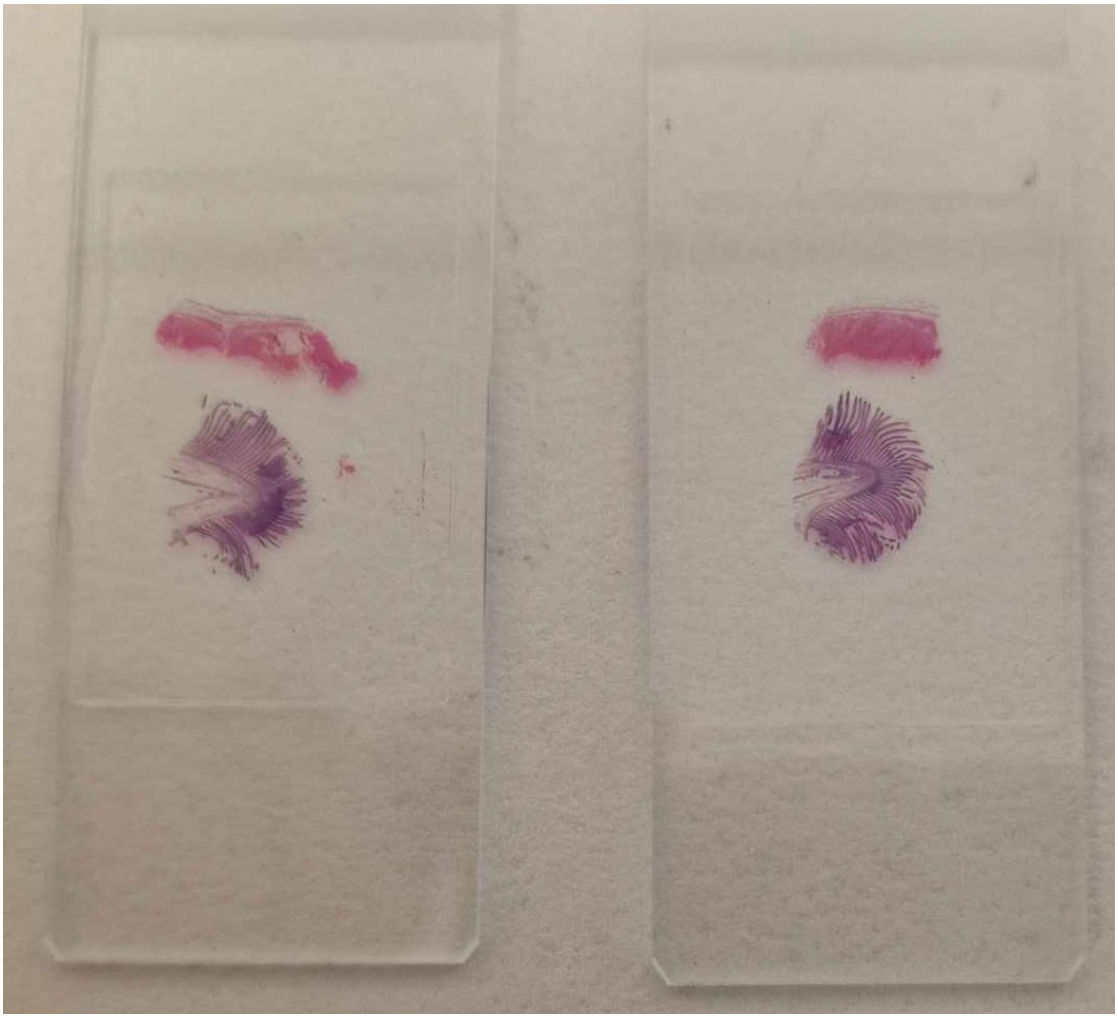


Figure 2.5: Completed gill slides after H&E staining. Gill tissue is fixed in formulin before being sectioned, then mounted on glass slides before going through the staining process.

7. **Image Capture**: Modern histology labs often use digital imaging systems to capture high-resolution images of the stained tissue sections. These digital images can be stored, analysed, and shared electronically. Figure 2.6 shows a Zeiss Axio Scan.Z1 as an example of these digital imaging systems.

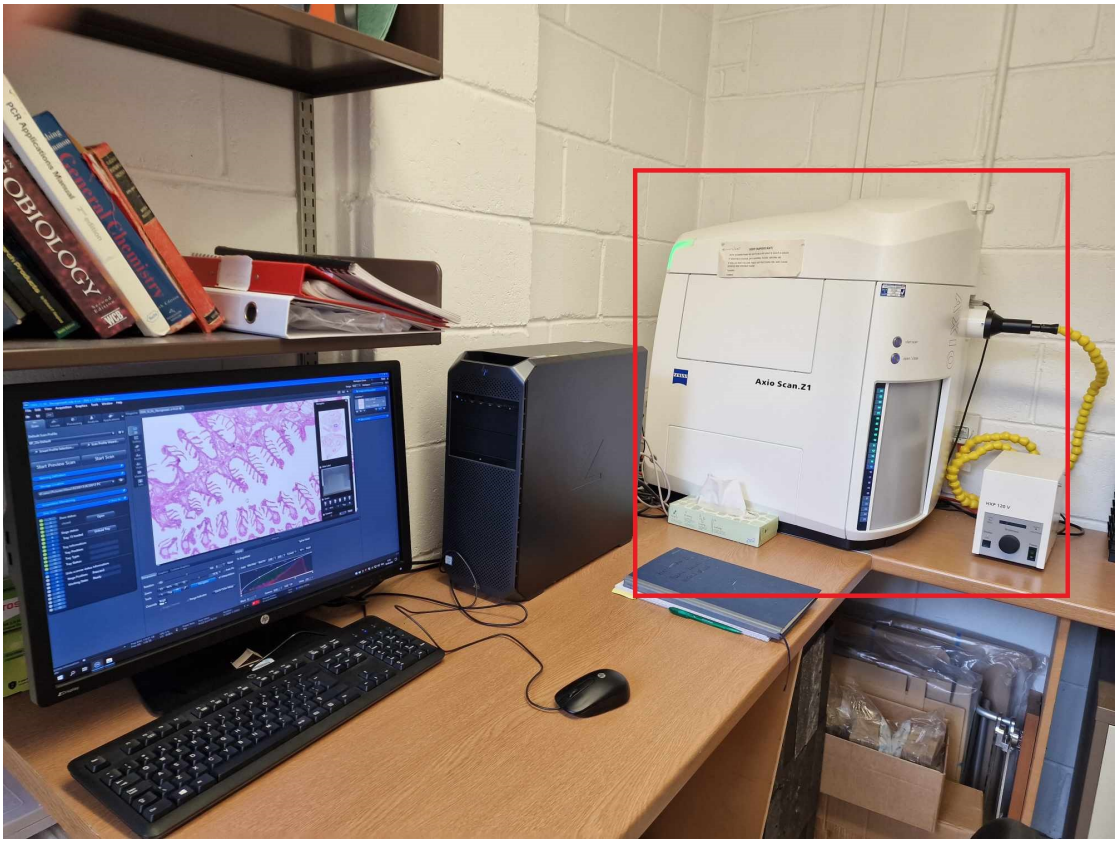


Figure 2.6: Zeiss Axio Scan.Z1 imaging system and a display showing a gill WSI.

Gill Tissue Morphology

Surface gill tissue is predominantly made up of branches, known as lamellae. An example of a gill whole slide image (WSI) can be seen in Figure 2.7.

The purpose of the lamellae is to increase surface area to aid in respiration. Each large branch, known as the 'Primary' lamellae, also has many smaller 'Secondary' lamellae branching from it. This can be seen in the image in Figure 2.8.

Secondary lamellae are composed of capillaries surrounded by a thin layer of epithelium. Mucous cells can also be found on the surface of the lamellae. Figure 2.9 shows examples of the cells and tissues described.

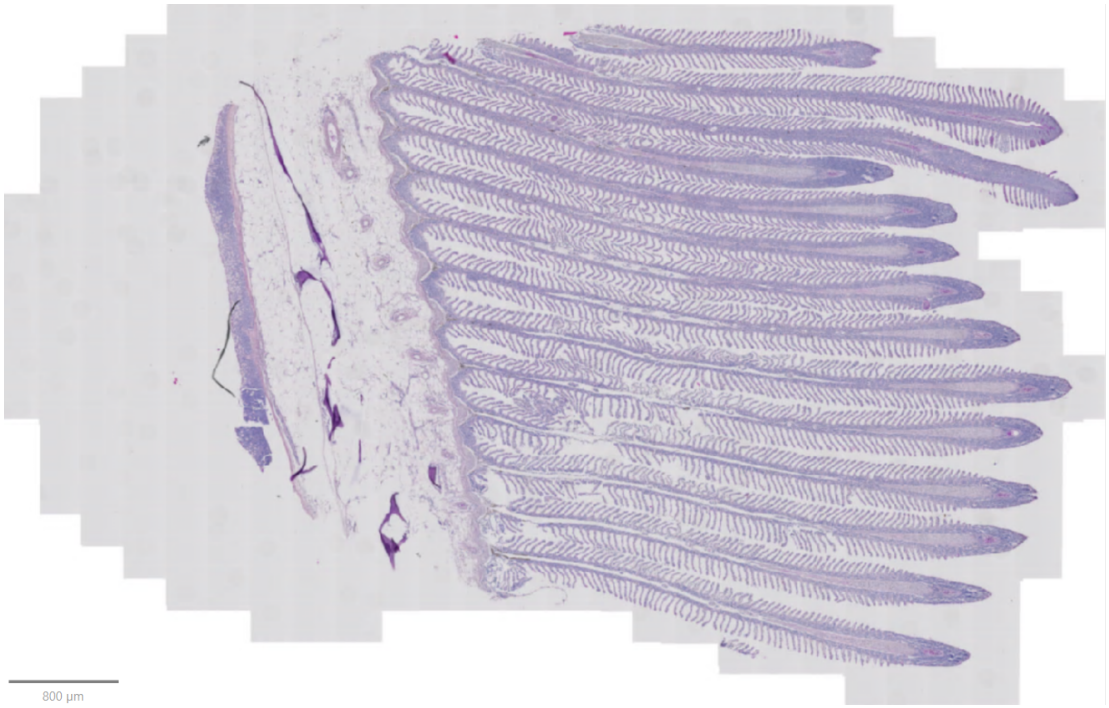


Figure 2.7: An example of a gill tissue slide whole slide image. The tissue has been stained using H&E staining. This example is a relatively healthy gill.

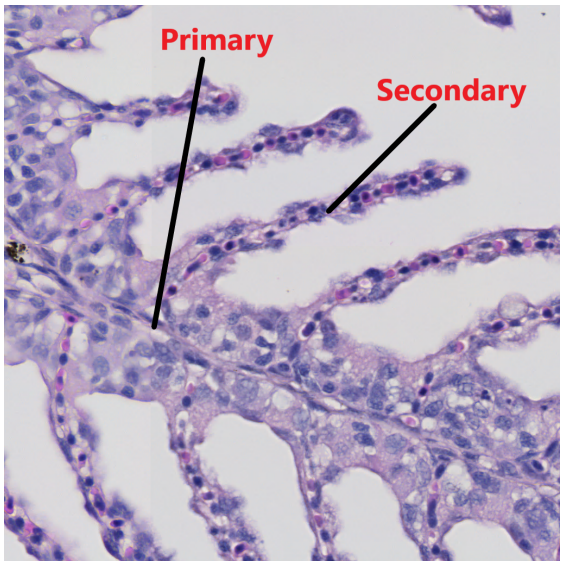


Figure 2.8: A labelled image showing the morphology of the primary and secondary lamellae. This tile is 1024x1024 pixels in size.

Current Applications of Data Science in Aquatic Animal Health

In a recent study by Gladju et al. in 2022, the authors conducted a comprehensive review of the applications of data mining and machine learning

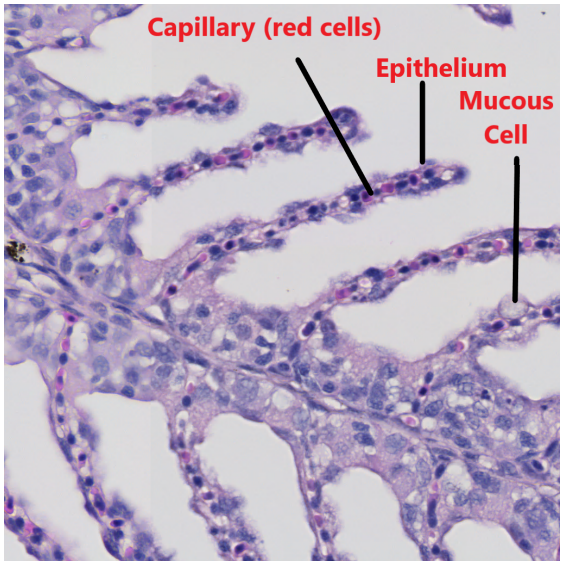


Figure 2.9: A 1024x1024px labelled image showing some of the cells and tissues present on the lamellae. Red cells are surrounded by a thin epithelial layer. Other cell types are also common, such as mucous cells.

techniques in aquaculture and fisheries. One of the key findings was the use of automation to enhance production environment monitoring, focusing on water quality parameters like temperature, pH, water levels, and overall water and waste management. Additionally, the review identified other valuable applications, such as optimising feed utilisation and assessing stock using various structured parameters, including gender, species, biomass, and product quality. Notably, the study by Rohani et al. in 2019 [30] was highlighted, where they employed traditional image processing features such as GLCM and utilised support vector machine (SVM) and multi-layer perceptron classifiers to separate live and dead rainbow trout fish eggs intelligently. However, it's worth noting that this method required the eggs to be taken out of the water for analysis, indicating the early stage of many machine-learning applications in aquaculture.

An earlier review in 2016 by Saberioon et al. [31] specifically delved into vision applications in aquaculture. Much of their review centred on utilising computer vision methods for behaviour analysis, employing various sensors, including acoustic, stereovision, and LiDAR. They also explored imaging applications for tasks such as fish sorting and assessing fish quality and attributes using X-ray, thermal, and hyperspectral imaging systems.

Since the publication of these reviews, more advanced deep-learning tech-

niques have gained popularity. For example, Fernandes et al. in 2020, [32] applied deep learning-based image segmentation to estimate the size and weight of tilapia carcasses. Many subsequent reviews have emphasised the wide range of computer vision applications in fish behaviour analysis. However, it is noteworthy that none of the reviews available at the time of this writing has highlighted the potential impact of applying machine learning methods to aquatic animal medical images [33–38]. This omission is surprising, considering the significant advancements in human medical imaging, some of which we discuss in Chapter 3.

In Chapter 3, we explore some parallel advancements in medical imaging, which could potentially offer valuable insights for applying machine learning methods to aquatic animal medical images. As technology continues to evolve, there is a compelling opportunity for cross-disciplinary collaboration and knowledge transfer between human and aquatic animal medical imaging. Such collaboration could usher in a new era of innovation, unlocking novel solutions to challenges in aquaculture and fisheries and ultimately contributing to the sustainable growth of these vital industries.

2.2 Technical Background

This section serves as a technical foundation for the thesis, offering an overview of fundamental concepts which will be referred to later in the thesis, laying the groundwork for understanding the main ideas and methods used in the research.

2.2.1 Mining Free Text

Extracting useful and meaningful information from free text can be done in many ways. These techniques can involve basic analysis, such as word frequencies and collocations.

- *Word/Term Frequencies* Identifying the most commonly used terms across a text or texts. This can be done at the unigram, bigram, trigram or n-gram level.

- *Collocation* Identifying sequences of words frequently appearing near each other. A common method for this task is Pointwise Mutual Information or PMI score. [39] The formula for the PMI score is defined in Equation 2.1, where the PMI score is the log probability of words co-occurring divided by the product of the single probabilities of each individual word occurring.

$$pmi(x;y) \equiv \log \frac{p(x,y)}{p(x)p(y)} \tag{2.1}$$

More complex methods may involve text classification tasks, including topic modelling, intent detection and sentiment analysis. Text extraction methods are particularly important to this work, involving keyword extraction, feature extraction and named entity recognition. [39]

Text Extraction

Common methods of extracting information from text include regular expressions, part-of-speech tagging, chunking and negation detection [39].

- *Regular expressions*- A syntax that allows one to define pattern rules for searching in text.
- *Part-of-speech tagging*- A process involving tagging each word in a text with its relevant part-of-speech.
- *chunking* Grouping part-of-speech tagged words based on a predefined grammar. The grammar can incorporate regular expressions to make complex pattern rules.
- *Negation detection* Establishing if a term used in a sentence has been negated based on the context of its usage.

2.2.2 Empirical Wavelet Transform

In traditional wavelet-based approaches for image decomposition, the image is processed through a series of filters, typically a combination of high-pass and low-pass filters, to create a set of subband images. These subband images represent different frequency components of the original image. The process is iteratively repeated by applying additional high-pass or

low-pass filters to these subband images until the desired representation is achieved. This approach has been widely used for various image-processing tasks.

In 2014, Gilles et al. revisited and enhanced several well-known traditional 2D transforms commonly used in image processing. Their work proposed empirical versions of four important transforms: Curvelets, Ridgelets, the Tensor approach, and the 2D Littlewood-Paley (LP) transform [40].

The traditional 2D Littlewood-Paley (LP) transform filters images in the Fourier domain using 2D wavelets with annuli supports centred around the origin. These annuli supports serve as the basis for dividing the scales, and their radii are determined based on the dyadic decomposition of the Fourier plane. Gilles et al. [40] proposed an empirical approach to detect the radii of these annuli.

In our methodology, we employ the 2D LP-EWT (Littlewood-Paley Empirical Wavelet Transform), denoted as $\mathcal{W}^{\mathcal{ELP}}$, to generate subband images from the input image f. The following equation describes the 2D LP-EWT:

$$\mathcal{W}_f^{\mathcal{ELP}}(n, \mathbf{x}) = \mathcal{F}_2^* \left(\mathcal{F}_2(f)(\omega) \overline{\mathcal{F}_2 \left(\psi_n \right)(\omega)} \right). \tag{2.2}$$

where $\mathcal{F}2$ denotes the 2D Fourier boundary detection function, $\mathcal{F}2^*$ represents its inverse, ψ represents the wave function, $\mathbf{x} = (x_1, x_2)$ represents the position in the 2D plane, and $\omega = (\omega_1, \omega_2)$ represents the position in the frequency domain.

We empirically choose $n = 20$ in our specific implementation, generating 20 subbands. The central panel of Figure 6.3 illustrates the frequency domain image, the boundaries generated by the 2D LP-EWT, and the resulting subband images.

The boundary detection method used to construct the filter bank, denoted as $\mathcal{B}$, plays a crucial role in the LP-EWT approach. To achieve this, we employ a Pseudo-Polar Fast Fourier Transform (FFT), denoted as $\widetilde{\mathcal{F}}_P(|\omega|)$, on the original image, taking into account the specified number of filters.

The first step in the boundary detection process involves performing the Pseudo-Polar FFT on the image, resulting in a transformed spectrum $\widetilde{\mathcal{F}}_P(|\omega|)$. This transformation is performed in a manner that emphasises the radial

frequency components.

Next, we calculate the average within each discrete angle θ in the 1D Fourier spectrum, where the number of angles is determined by N_θ. This averaging process is carried out to capture the information across different angles in the frequency domain.

Mathematically, this process can be described as follows:

$$\widetilde{\mathcal{F}}_P(|\omega|) = \frac{1}{N_\theta} \sum_{i=0}^{N_\theta - 1} \left| \mathcal{F}_P(f)\left(\theta_i, |\omega|\right) \right|. \tag{2.3}$$

where $\mathcal{F}\left(|\omega|, \theta\right)$ represents the 2D Fourier transform of the image at the radial frequency $|\omega|$ and angle θ.

By applying this boundary detection method, we effectively capture the relevant frequency information across different angles, which is essential for constructing the filter bank $\mathcal{B}$ used in the subsequent steps of the LP-EWT. This process allows us to extract meaningful features from the image that can further contribute to the accurate classification and analysis of hyperplasia in Atlantic Salmon gill whole-slide images (WSIs).

The filter bank is then constructed using the spectral radii obtained, shown in Eq. (2.4). More details can be found in Gilles et al.'s 2014 paper on 2D empirical wavelets [40]. The outcome of this process are the boundaries shown in the central panel of Fig. 6.3.

$$\mathcal{B} = \left\{ \phi_1(\mathbf{x}), \{\psi_n(\mathbf{x})\}_{n=1}^{N-1} \right\}. \tag{2.4}$$

3 State-of-the-Art

This chapter explores state-of-the-art information retrieval and data mining applications in relevant medical texts, histology image processing, empirical wavelet transforms, and artificial intelligence (AI) in medical imaging. Our objective is to provide a clear overview of recent advancements and trends in these specific domains, which form the foundation of our research. We will critically examine existing literature, highlighting the gaps and opportunities that have informed our research agenda. This contextualisation will set the stage for a detailed examination of our own research contributions and their significance within these specialised fields.

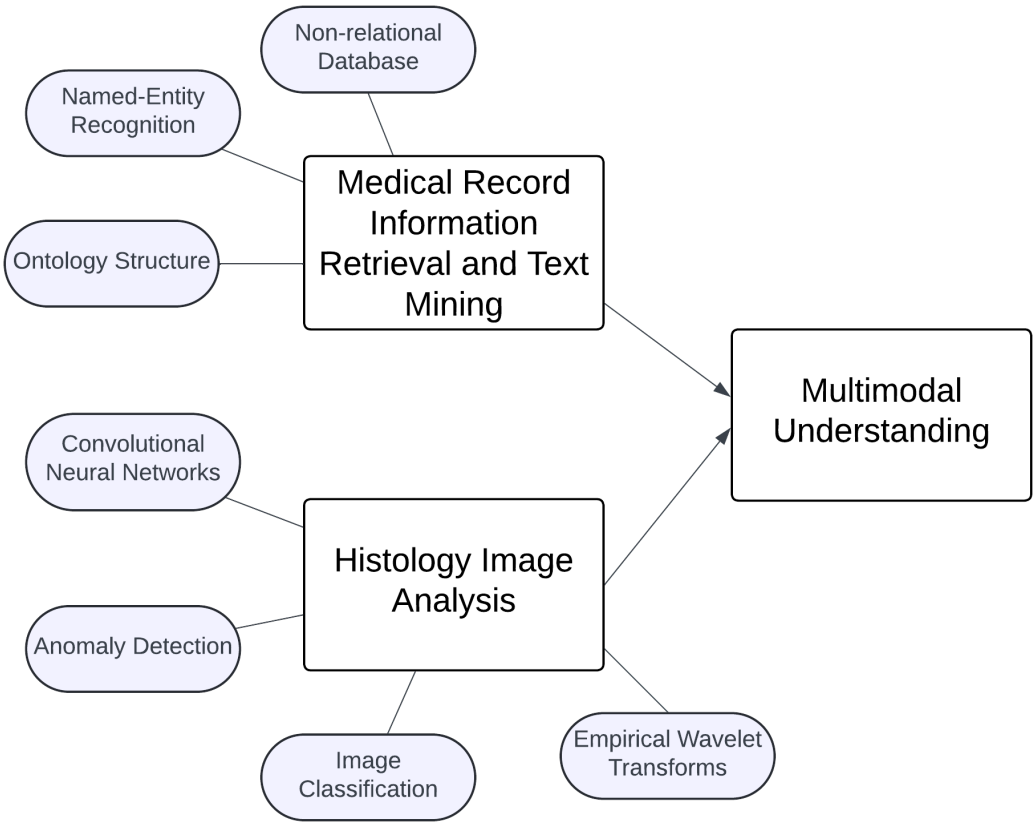


Figure 3.1: A diagram which describes the relationships between different components of the project. Technology components, highlighted in blue, are associated with tww main tasks in this project, which can both contribute to multimodal understanding of cases.

3.1 Medical Record Information Retrieval

Information retrieval is important for automated data extraction from free-text medical records. Various methods exist, mostly consisting of text mining and text classification approaches. While current literature extensively covers information retrieval approaches in the biomedical domain (particularly on human health), fewer attempts have been made in the veterinary domain.

Related approaches for information extraction in the biomedical domain are numerous. Chapman *et al.* [41] proposed NegEx, a tool for determining the presence or absence of clinical findings in discharge summaries. Their

approach was used to analyse $76,049$ screening and $17,656$ diagnostic mammography reports. Even though this approach extracts conclusions rather than observations – which are the focus of our work - the applications are similar. More recently, Gao *et al.* [42] extracted several features from mammographic reports: mass, calcification, asymmetry and architectural distortion.

Friedlin *et al.* [43] developed Medical Exploratory Data Analysis over Text (MEDAT), a text analytics system for medical domains and demonstrated their system on radiology reports. Comelli *et al.* [44] also applied text mining to radiology reports. They leveraged the entity relationships represented in their radiology ontology to extract relevant medical terms from mammographic reports. While their approach was exhaustive within the mammographic domain, the texts were in Italian, and the application domain is quite narrow.

Gong *et al.* [45] developed a biomedical information retrieval approach for terminologies related to breast cancer. Their approach involved entity extraction, entity relationship identification, and visualisation. Entities were extracted based on conditional random fields, while co-occurrence statistics extracted entity relationships. Sudeshna *et al.* [46] aimed to identify symptoms and treatments of heart disease using a machine learning-based approach. Based on suggested identified symptoms, texts would be classified into treatments. Zhao *et al.* [47] created CausalTriad, an approach toward discovering pseudo-causal relationships between entities. They evaluated their approach on HealthBoards message board and Traditional Chinese Medicine data. Yang *et al.* [48] used an ontology-based text mining approach to extract data from Chinese Electronic Medical Records (EMR). This work focused on the mining of stroke cases. Gero and Ho [49] proposed NamedKeys, a keyphrase extraction approach which they evaluated on PubMed abstracts. They also describe a benchmark dataset for biomedical keyphrase extraction.

While various clinical text types have been the subject of much research, there is also a wide body of research into automated biomedical literature reviews. Navathe [50] used UMLS (Unified Medical Language System) [51] and a gene ontology to represent biomedical concepts, and a Support Vector Machine (SVM) to classify literature from the Centre of Disease Control

(CDC) based on relevant keywords. Mala *et al.* [52] researched the use of ontology in semantic medical text mining with WordNet. Gong *et al.* [53] used a dictionary-based approach to extract biomedical concepts from literature. This was done using an algorithm called the Variable-step Window Identification Algorithm (VWIA), which matched terms to biomedical entities using POS tagging and organisation based on phrasing. Their technique was applied to 10 Medline abstracts and produced promising results. Mate *et al.* [54] focused on creating a process of extraction, transformation and loading (ETL) of electronic medical records.

Although not used in the approach described in Chapter 4, it should be noted that emerging *deep learning* has become popular in the biomedical domain with neural network-based methods being used to enhance text mining techniques [55, 56].

All of the reports listed above applied information retrieval techniques to biomedical text about humans. In the veterinary domain, Bollig *et al.* [57] used machine learning-based approaches to extract different pathologies from free text. Furrer *et al.* [58] built a text mining tool for veterinary surveillance by linking terms identified in necropsies to existing ontologies. Küker *et al.* [59] later used this tool to analyse pig and cattle necropsies and found that free text necropsy reports are a valuable resource for animal health surveillance.

At the beginning of this project, no work existed on information retrieval from marine mammal necropsy reports.

In recent years, natural language processing has witnessed a monumental shift marked by the emergence of robust large language models. Among these, ChatGPT, introduced by OpenAI in 2022 [60] as an evolution of GPT-3.5 [61], stands as a prominent example. This pivotal development set the stage for a rapid succession of similar models, including Bard by Google [62] and LLaMA by Meta [63], among others.

This disruptive technology has sparked extensive discussions regarding its potential impact across various domains, with medicine being a particularly noteworthy area of interest. For instance, Liévin et al. demonstrated that GPT-3.5 achieved human-level performance on three biomedical research question-answering datasets [64], highlighting its promising appli-

cations in biomedicine.

3.2 Multimodal Vision and Language Tasks

In recent years, machine learning techniques have widely applied to handling text and image data separately. However, integrating both modalities in tandem has received considerably less attention. Multimodal sentiment analysis, nevertheless, represents a promising field where such integration has been explored. Research in this domain primarily focuses on three key areas: the analysis of vlogs and spoken reviews, the examination of interactions between humans and machines, as well as computer-computer interactions, and the analysis of tagged images sourced from social media platforms [65].

Document classification has benefited significantly from multimodal analysis, with recent approaches demonstrating its potential [66–68]. These methods leverage Optical Character Recognition (OCR) to extract text from document images, utilising both the visual content and the in-document text to classify documents into various types. Many of these approaches rely on open-source OCR software like the Tesseract OCR library [69] for accurate text extraction.

The methodologies employed in this task could potentially find application in processing histology images and reports. However, the necessity for OCR methods would depend on the accessibility of textual information within the histology data.

In the biomedical field, Pelka et al. introduced the Radiology Objects in Context (ROCO) dataset [70]. This dataset includes diverse medical imaging modalities, captions, keywords, and tags. While ROCO has predominantly served as a resource for tasks such as image tagging and concept detection [71–73], its richness, encompassing multiple textual modalities, also positions it for potential applications in multimodal visual and textual classification.

3.3 Histology Image Analysis

Numerous organisations and stakeholders within the aquaculture industry accumulate a substantial volume of histology Whole Slide Images (WSI). However, it is noteworthy that while an extensive body of research focuses on automated analysis of human anatomical specimens, comparatively little attention has been directed towards the analysis of fish. Developing specialised and efficient techniques for analysing gill WSI can significantly enhance the efficiency and depth of analysis for conditions like Complex Gill Disease (CGD) in fish populations.

A diverse range of approaches exists for WSI analysis, spanning both traditional image processing techniques and neural network (NN)-based methods. In non-NN approaches, the preprocessing of WSIs usually encompasses image transformations and extracting relevant features. In contrast, many deep learning-based methods leverage convolutional and pooling layers within neural networks for comprehensive image analysis.

In histology, tissue samples are often treated with dyes, a common practice being the application of Hematoxylin & Eosin (H&E) stain, which is the primary focus of our work. Stain normalisation plays a pivotal role in most analysis pipelines. It addresses the challenge of inconsistent colour intensities between WSIs due to variations in stain application. Such inconsistencies can hinder the effectiveness of image transformation and feature extraction techniques. Stain normalisation techniques like Vahadane [74], Reinhard [75], and Macenko [76] are commonly employed to mitigate these issues.

In the realm of human anatomy, pathologies such as breast cancer [77], lung cancer [78–81], and brain tumors [82] have garnered significant research interest. Applications encompass carcinoma classification, blood cell detection [83], nuclei detection, and cell segmentation. In recent years, Convolutional Neural Networks (CNNs) have gained popularity as a prevailing approach for tackling these tasks [29, 84–86].

The surge in popularity of Convolutional Neural Networks (CNNs) can be partially attributed to the substantial advancements in hardware performance over recent years. This progress has rendered training machine

learning models with large datasets a practical and effective strategy. CNNs, equipped with their convolutional and pooling layers, have become the preferred choice over traditional feature extraction methods reliant on statistical colour and texture representations. The adaptability and efficiency of CNNs in handling complex visual data have made them a formidable tool in modern machine-learning applications.

Despite the widespread adoption of CNNs in WSI analysis, traditional texture metrics maintain their prominence and effectiveness as valuable means of feature extraction. Notable texture feature sets applied in WSI analysis encompass local binary patterns (LBP), Haralick's texture features (derived from a grey level co-occurrence matrix or GLCM), and Gabor filters [82]. Additionally, both lower-order and higher-order histograms have found utility in this context.

An illustrative study by Kather et al. [87] underscores the continued relevance of texture analysis techniques. In their investigation of colorectal cancer classification, they discovered that lower-order texture measures, GLCM, and LBP emerged as the most effective texture representations for their specific use case, showcasing the enduring value of these methods in certain medical imaging applications.

Machine learning (ML) has found increasing application in aquaculture to optimise fish growth and performance, as noted by Zhao et al. [88]. Surprisingly, there has been limited development in automated histological analysis of fish gill tissue despite its critical role in the field. Existing methods have predominantly relied on traditional image preprocessing steps and alternative staining techniques, such as quantifying mucous cells in salmon skin [89] or assessing gill health [90]. Jayasuriya introduced a tool for evaluating morphological changes in salmon gills, generating descriptors automatically individually [91]. While beneficial for in-depth analysis of single gill WSIs, there is a growing need for data-driven approaches capable of efficiently handling large numbers of WSIs.

Developing template-based computer vision pipelines for such tasks can be challenging, prompting interest in data-driven machine learning-based approaches. Sveen et al. successfully applied deep learning to automatically segment Atlantic Salmon skin tissue [92], but there has been limited progress in automating the analysis of gill histology images. Notably, Swee

et al. [93, 94] used coloured histogram and Gabor wavelet transform features in conjunction with Support Vector Machines (SVM) to detect water pollution, albeit on WSIs of Tilapia liver samples. Similarly, Silva et al. [95] employed image processing and segmentation techniques to assess Atlantic Salmon intestine morphology, including smoothing, noise removal, edge detection, thresholding, and region filling. Although their work was limited to the intestines, similar approaches could be applied to other fish tissues.

Considering the dearth of research in gill histology image processing, the development of a method for automated analysis of WSIs holds significant promise and potential impact.

3.4 Empirical Wavelet Transform for Image Analysis

Automating biomedical image classification tasks can be achieved through frequency domain texture analysis. However, traditional image decomposition methods typically fall into two categories: fixed template-based approaches (e.g., wavelets) or data-driven methods (e.g., Empirical Mode Decomposition or EMD [96]). The former can lack flexibility due to their reliance on rigid template structures, while the latter may lack a well-defined mathematical foundation. In an effort to bridge this gap, a recent development by Gil and others [40, 97] introduced the empirical wavelet transform (EWT), which offers adaptability to the input signal, providing a more versatile approach to texture analysis in the frequency domain.

Traditional wavelet-based techniques typically entail the decomposition of an image into a collection of subband images by applying various filters, typically a combination of high-pass and low-pass filters. In 2014, Gilles and colleagues revisited several well-known traditional 2D transforms, introducing empirical variants of Curvelets, Ridgelets, the Tensor approach, and the specific method employed in this study—the 2D Littlewood-Paley (LP) transform [40]. This development aimed to enhance the adaptability and performance of these transforms in image analysis tasks.

The EWT has found application in glaucoma classification using Fundus

Imaging, although it's important to note that the morphological structure of glaucoma cases tends to exhibit more consistency compared to the variability seen in gill Whole Slide Images (WSIs). In gill WSIs, primary and secondary lamellae can vary in structure based on their location, and overlapping branches can further complicate the analysis. Moreover, the tissue sectioning process often introduces irrelevant artefacts.

In the context of glaucoma classification, Maheshwari et al. utilised 2-D empirical Littlewood-Paley (LP) wavelet subband images to extract correntropy features for the identification of glaucoma cases in fundus images [98]. Similarly, Kirar et al. developed an EWT-based approach that combined traditional discrete wavelet transforms with 2D LP-EWT to derive features from subband images, including Zernike moments, Hu's Invariant Moments, chip histogram, and grey-level matrix approaches [99]. Chaudhary et al. [100] applied a Fourier-Bessel series expansion-based EWT (2D-FBSE-EWT) and transfer learning using pre-trained ResNet-50 models to evaluate subband images. Additionally, they employed their 2D-FBSE-EWT technique for denoising biomedical images [101]. These studies showcase the adaptability of EWT in different image analysis tasks, albeit in contexts with different structural characteristics compared to gill WSIs.

Other notable imaging applications of EWT include gastrointestinal abnormality analysis in endoscopic images [102], oral cancer histology image classification [103], COVID-19 detection using chest CT images [104], lung disease diagnosis [105], and melanoma detection in skin images [106]. These studies demonstrate image analysis techniques' versatility and wide-ranging impact in diverse medical contexts.

These approaches utilise traditional feature extraction techniques and CNNs to analyse the subband images generated by the EWT transform. While CNNs have gained popularity for various histology tasks, it's important to acknowledge their inherent weaknesses, which include their black-box nature and lengthy training times. These drawbacks can pose challenges in terms of interpretability and computational efficiency, particularly in the context of histology image analysis.

3.5 Anomaly Detection in Images

Anomaly detection involves the identification of outlier entries within a dataset [107]. This task has found application across various domains, particularly computer vision and medical imaging. In medical imaging, Alloqmani et al. [107] identified several notable use cases, such as detecting irregular tissue in mammographic images from the INbreast dataset [108, 109], screening for viral pneumonia in chest x-rays [110], tumor detection in brain MRIs [111], and anomaly detection in retina images [112, 113], among others.

Traditionally, image anomaly detection relied on "shallow approaches," which involved extracting features from images and passing them to a classifier. These methods often encompass techniques like principal component analysis [114], support vector machines (SVMs) [115], and nearest neighbour models [116]. In contrast, contemporary "deep" methods, including autoencoders [117], variational autoencoders (VAEs) [118], and generative adversarial networks (GANs) [119], have gained prominence. Deep anomaly detection methods operate on the principle that anomalies can be detected by assessing a model's ability to reconstruct an image accurately.

4 Information Retrieval for Marine Mammal Necropsy Analysis

4.1 Overview

In this chapter, I describe Ir-Man (Information Retrieval for Marine Animal Necropsies), an advanced framework designed to extract information from post-mortem reports concerning marine mammals [3]. When a deceased marine mammal is discovered stranded along the shores of Scotland, the Scottish Marine Animal Strandings Scheme (SMASS) conducts a thorough examination to investigate and determine the cause of death. A comprehensive "post-mortem" or necropsy report is generated as part of this examination, providing a detailed account of the carcass, including observations of lesions and other relevant findings. These observations play a critical role in identifying the cause of death and may even offer significant insights from a pathological and epidemiological standpoint when analysed collectively.

Unfortunately, manually extracting this vital information from the reports is arduous, time-consuming, and error-prone. Consequently, there exists a need for an automated information retrieval system that can effectively handle the diverse array of terms encountered in these reports. The Ir-Man framework has been developed to address this challenge, encompassing three key components: an ontology, a lexicon of observations and anatomical terms, and an entity relation engine.

The ontology component is a structured framework that defines and organises the domain-specific knowledge required for information retrieval from necropsy reports. It establishes a systematic and consistent representation of the concepts and relationships related to marine animal necrop-

sies, enabling efficient data retrieval and analysis.

The lexicon of observations and anatomical terms forms an extensive collection of terminology relevant to marine mammal pathology. It encompasses a wide range of descriptors, ensuring that the Ir-Man framework can effectively identify and extract specific information from the reports, regardless of the variations in descriptions, pathologies, and species encountered.

The entity relation engine, a vital component of the Ir-Man framework, enables extracting and correlating information from a pool of necropsy reports. By leveraging advanced techniques and algorithms, this engine identifies the relevant entities and their relationships within the reports, facilitating comprehensive information retrieval and statistical analysis.

To showcase the effectiveness of the Ir-Man framework, we conducted a specific case study focusing on identifying bottlenose dolphin attacks (BDA) in gross pathology reports of harbour porpoises. Using a rule-based binary classifier developed within the Ir-Man framework, we achieved an accuracy rate of 83.4% in detecting BDA cases.

4.2 Framework Structure

While developing the Ir-Man framework, we carefully considered several steps involved in extracting observations from marine mammal necropsies. These steps ensure a systematic and accurate information retrieval from the necropsy documents.

The first step involves extracting the free text from the necropsy documents. This includes pulling the relevant sections, such as the gross pathology report section and, if applicable, the histopathology and bacteriology report sections.

Once the text is extracted, it is further processed by dividing it into sentence and word-level tokens. This division enables us to analyse the text at a granular level and facilitates subsequent tagging of individual words based on their part of speech. The tagging process helps identify the role and function of each word within the context of the text.

To establish meaningful relationships between entities, we employ a feature-

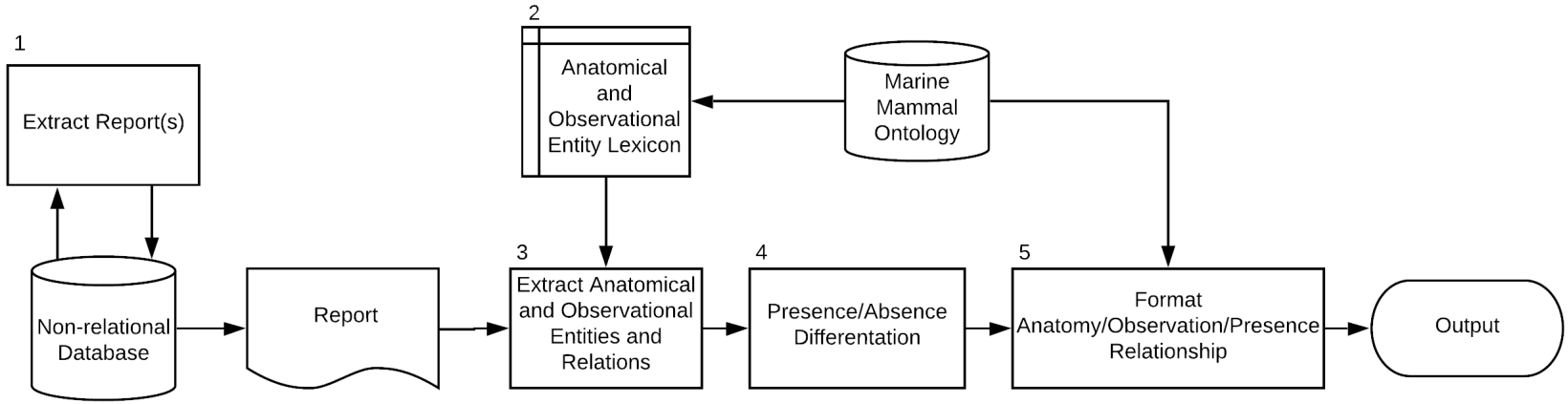


Figure 4.1: Overview of the Ir-Man framework along with its information retrieval flow diagram. The diagram shows how information is extracted using our ontology driven approach. Entity relations are extracted using the lexicon developed from the marine mammal ontology.

based grammar approach. This technique groups related entities and captures the connections and associations between them. By structuring the information based on these relationships, we enhance the coherence and organisation of the retrieved data.

Next, each identified entity is compared against our comprehensive anatomical, pathological, and observational lexicon. This lexicon is generated using our ontology, which provides a structured framework for organising and categorising the relevant domain-specific knowledge. By leveraging this lexicon, we ensure the accuracy and consistency of the extracted information.

Considering the presence and absence of described features in the retrieval process is important. Thus, we explicitly record negative occurrences or absences of identifiers in addition to positive occurrences or presences. This comprehensive approach allows us to leverage both types of information in developing a deterministic classification system.

To provide a visual overview of the retrieval process, we outline it in Figure 4.1. The accompanying pseudo-code in Algorithm 1 further illustrates the implementation steps. Each component of the Ir-Man framework plays a crucial role in achieving accurate information retrieval, and we provide detailed descriptions of these components below.

By meticulously considering and implementing these steps, the Ir-Man framework ensures a robust and reliable process for extracting and organising observations from marine mammal necropsies, enabling comprehensive analysis and valuable insights.

4.2.1 Marine Mammal Stranding Reports

The data utilised in this project was derived from post-mortem (PM) reports of cetaceans, specifically generated by the Scottish Marine Animal Stranding Scheme (SMASS) from 2012 to 2019. These reports document various features of the carcass, including its condition, morphology, pathological lesions, and observations. Assessing the body's condition typically encompasses factors such as the extent of autolysis or physical damage. Additionally, morphometric measurements, such as blubber thickness and body length, are recorded. These features contribute to the pathologist's under-

Algorithm 1: Information retrieval pipeline. Output of the entity-relationship extraction engine is used to identify observations, attributed anatomies and detect negation.

```
Result: relationships
sentences = sentenceTokenisation(text);
observations;
while not at end of sentences do
    RELChunkedSentence = preprocess(sentence);
    identifyNamedEntities(RELChunkedSentence);
    while not at end of sentences do
        if No Observational Entities then
            break to next relationship;
        end
        if Observational Entity AND No Anatomical Entity then
            observations <- 'unattributed' observation;
            break to next relationship
        end
        if Observational Entity and Anatomical Entity then
            observations <- anatomy, observation;
        end
    end
end
while not at end of observations do
    negatedObservation <- mark_negated(observation) ;
    if observation == negatedObservation then
        presence <- True;
    else
        presence <- False;
    end
end
```

standing of the probable cause of death. PM reports consist of detailed gross pathology reports, providing comprehensive descriptions of the overall characteristics of the body, as well as specific anatomical details. The final PM report comprises several sections, including basic information (such as sex, date, and location), morphometric data, gross pathology findings, bacteriology and histopathology reports (if applicable), and a conclusion that encompasses comments, cause of death, and the level of confidence in the diagnosis.

For this study, we analysed a dataset consisting of 193 gross pathology reports specifically focusing on harbour porpoises (*Phocoena phocoena*). This

species was chosen for several reasons: SMASS has produced a relatively high number of reports on harbour porpoises, making it an easily accessible dataset; the anatomy of the harbour porpoise is transferable to other cetacean species, enabling potential future integration of additional species; and the prevalence of bottlenose dolphin attacks (BDAs) listed as the cause of death, which allows us to evaluate the effectiveness of our framework in detecting the pathologies associated with such attacks. Bottlenose dolphins exhibit violent behaviour towards harbour porpoises, often leaving distinct parallel incisions known as "rake marks." These rake marks serve as a primary indicator of a BDA, and the consistent usage of this term facilitates the evaluation of our approach's efficacy.

Although the language employed in these gross pathology reports is specialised, a certain degree of structure can be leveraged. Typically, a heading indicating the anatomical region of interest is followed by a description in free-text format. This structure can be observed in Figure 4.2. Furthermore, acronyms such as NAD (no abnormalities detected) and NE (not examined) hold particular significance and are distinct in their meaning. The absence of abnormalities can rule out certain pathological conditions, while the lack of examination in a specific region does not provide the same level of information.

By considering these unique characteristics and leveraging the structured elements within the reports, our framework aims to effectively extract and analyse the specialised language used in gross pathology reports. This approach enables us to identify important anatomical regions and associated descriptions and interpret acronyms that carry specific diagnostic implications.

Body condition: Fat

External examination
Body orifices: NAD
Ectoparasites: NAD – None seen
Fins and flukes: NAD – Intact, no rake marks

Integument
Epidermis: Rake marks over left flank/tailstock. Severe scavenger damage at right side of head
Blubber: NAD – Good layer, not jaundiced
Subcutaneous tissue: Bruising over lateral spinous processes and right side head region
Mammary glands: NE

Figure 4.2: A gross pathology extract from a harbour porpoise necropsy report. Observations are attributed to anatomical regions. Accronyms such as NAD and NE are used to highlight when there are “no abnormalities detected", or if a region is “not examined".

4.2.2 Gross Pathology Report Extraction

The post-mortem reports from the Scottish Marine Animal Stranding Scheme (SMASS) were stored in Microsoft Word Open XML Format (DOCX) files. We parsed these documents to manage and access the data efficiently and stored the extracted fields in a non-relational MongoDB [1] database.

We implemented a search mechanism for specific text fields to identify field names indicative of their presence within the reports. For instance, when extracting the species field, we used the string “SPECIES:” as the indicator and captured the string following it on the same line as the corresponding field value (*e.g.*, “*delphinus delphis*”). In cases where a field was left blank, indicating no value, we did not store any information in the database for that field.

We applied normalisation techniques to the extracted fields to enhance data consistency and ease of analysis. Synonymous terms were grouped

[1] https://www.mongodb.com/

together to ensure standardisation. For example, in the case of the species field, we paired scientific names (*e.g.*, "*delphinus delphis*") with their corresponding common names (*e.g.*, "short-beaked common dolphin").

In addition to structured fields, we also obtained free text sections such as the gross pathology reports. We accomplished this by identifying relevant section headers and extracting the text situated between them. In cases where the space between section headers contained only white-space or short strings like "Not examined," we decided not to extract that particular section, as it lacked significant information.

All the extracted fields and sections were securely stored in a local MongoDB database. This database allowed for efficient storage, retrieval, and manipulation of the data, enabling further analysis and exploration of the post-mortem reports in a structured manner.

4.2.3 Ontology Development

Our framework utilises custom-built ontologies to organise and provide contextual information otherwise unavailable. While there are existing multi-species ontologies like Uberon [120] (a cross-species anatomy ontology categorising entities by traditional anatomical criteria), we opted for a smaller and more manageable ontology specifically tailored to our task. We identified three primary branches of relevant terminologies that align with our objectives.

The first branch focuses on representing anatomy. In this ontology, different anatomical regions are represented as classes. These classes allow us to categorise and identify specific areas of interest within the marine mammal carcasses.

The second branch is the pathology ontology, which serves as a means to record various conditions found in post-mortem (PM) reports. By referencing this ontology, we can systematically capture and categorise different pathological conditions described in the reports.

The third branch is the observation ontology. This ontology organises terms into classes and sub-classes, enabling a hierarchical observation structure. The sub-classes provide an additional level of specification that may

not apply universally to all entities within the parent class. This hierarchical organisation allows for a more precise and detailed representation of the observations made in the PM reports.

For all classes within the ontologies, we store representative labels in the “rdfs:label” annotation. Additionally, manually generated synonymous terms are stored in our own “synonym” annotation. These annotations facilitate better understanding and interpretation of the ontological terms and ensure consistency in data representation.

We utilised Protégé [121], a popular ontology development tool, to develop the ontologies. Figure 4.3 illustrates the visualisation of the ontologies within the Protégé interface. Once created, the ontologies are stored in the RDF/XML format, which allows for standardised representation and interoperability.

By employing these bespoke ontologies, our framework gains a structured foundation that enables effective organisation and classification of terms. The ontologies enhance the retrieval and analysis of information from the PM reports, leading to more accurate and meaningful results.

Observation Ontology

The observation branch of our ontology takes advantage of semantic relationships between terms to enhance the specificity and granularity of the data. When terms are semantically similar, but one provides a higher degree of specification, we establish a parent-child relationship between them. This hierarchical relationship allows for precise distinctions and categorisation of different types of observations in the post-mortem (PM) reports.

For example, consider the presence of fluid described in the reports. Some reports may simply mention “fluid,” while others may specify “brown fluid.” Since not all fluid is brown, we create a parent-child relationship between these terms. This relationship enables us to differentiate and classify various types of fluids and their specific descriptions, such as mucoid, protein-rich, or amniotic fluids.

To populate the observational ontology, we conducted a manual process. We generated lists of terms based on frequent unigrams, bigrams, and tri-

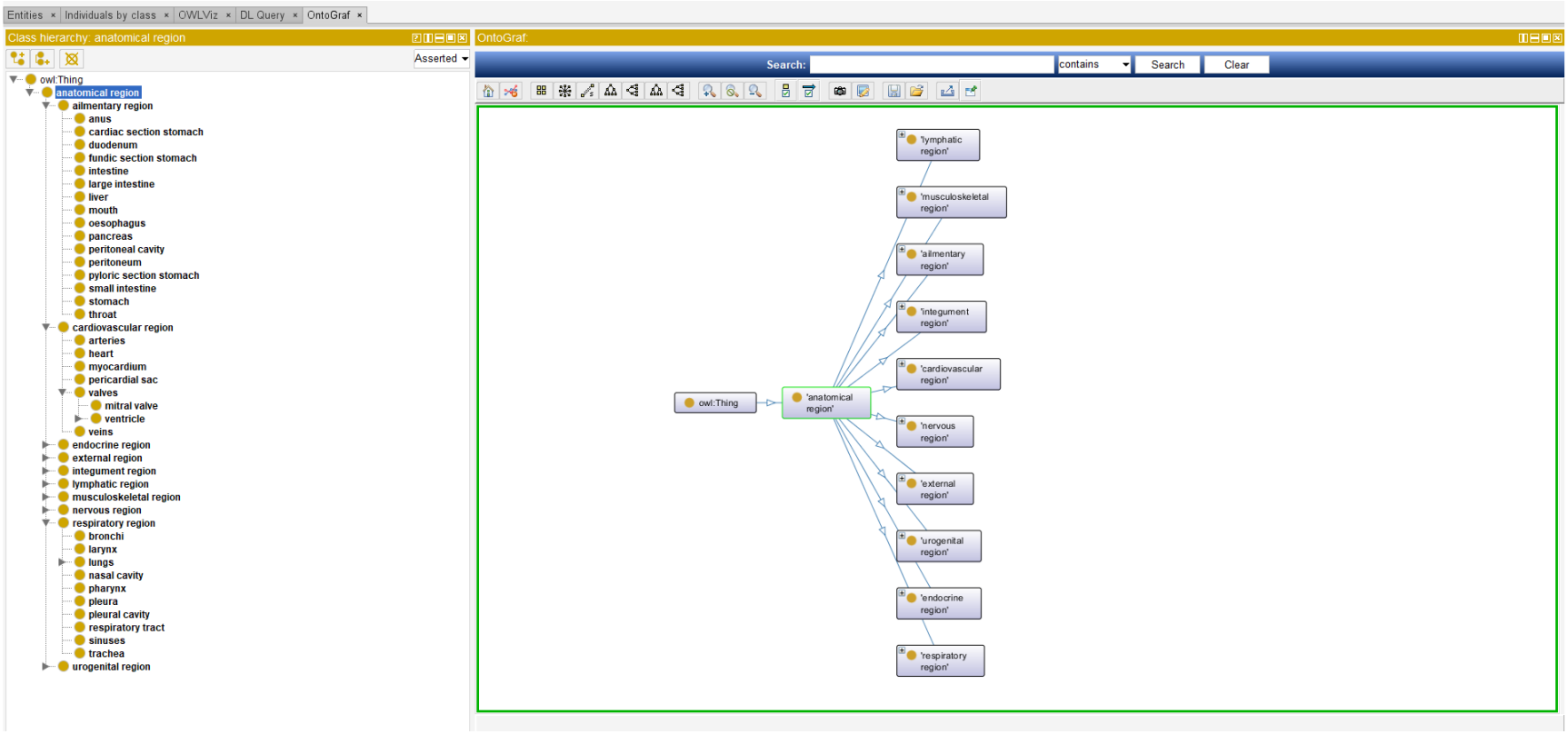


Figure 4.3: Screenshot of Protégé IDE for ontology development. The OntoGraf plugin [122] was used for ontology visualisation and development. This example shows how we developed the relationships between terms associated with different anatomical regions.

grams extracted from the reports. Additionally, we utilised collocations using Pointwise Mutual Information (PMI) to identify relevant terms. We filtered previously established anatomical and disease-related terminologies to expedite the process, focusing specifically on terms relevant to observations.

Figure 4.4 provides an overview of the structure of the observational ontology, illustrating the hierarchical relationships between different terms. This structure allows for a comprehensive representation of observations made in the PM reports, facilitating accurate classification and analysis of the data.

We ensure that the observational ontology captures the nuances and variations in the reports by employing semantic relationships and utilising statistical approaches such as pointwise mutual information (PMI) to identify salient terms. Our ontology enhances the precision and accuracy of our framework, enabling more robust retrieval and analysis of information from the PM reports.

Pathology Ontology

The pathology ontology, as depicted in Figure 4.5, plays a crucial role in representing different conditions and their semantic relationships. To construct this ontology, we initially utilised the diseases or conditions listed as causes of death within the SMASS database. Additionally, we extracted relevant information from the reports by identifying known target strings that commonly precede the causes of death. These extracted conditions were then categorised based on their semantic similarity.

For instance, within the ontology, we have a class named “physical trauma” that encompasses cases where there is evidence of blunt force or penetration to the skin, which have had a detrimental impact on the animal’s health. This category specifically includes conditions such as boat strikes, bottlenose dolphin attack trauma, and entanglement, where the animal has become wrapped in rope, line, or netting.

By organising these conditions into classes and defining their relationships within the ontology, we create a structured framework that compre-

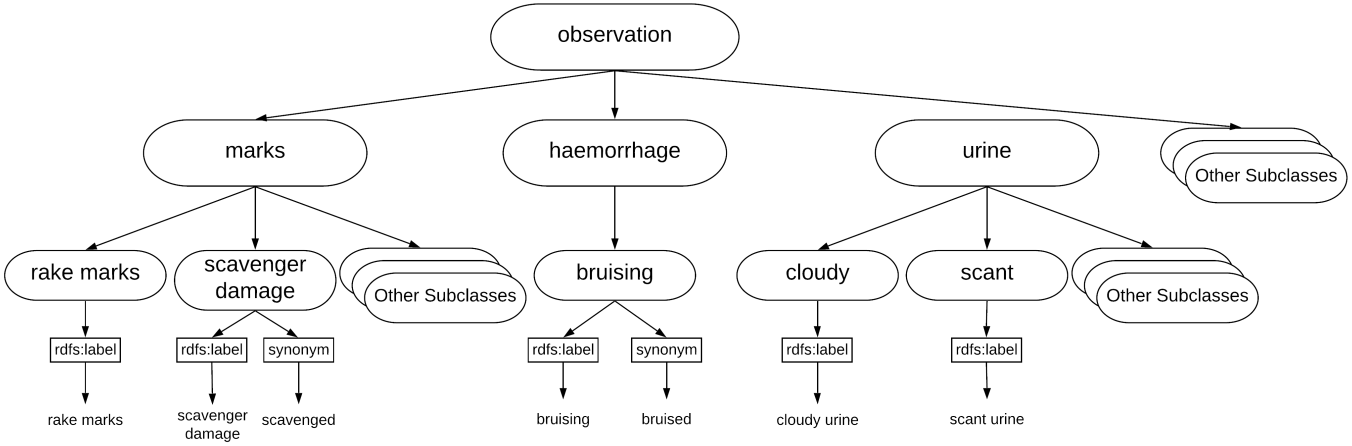


Figure 4.4: Structure of the observation ontology demonstrated using the marks, haemorrhage and urine class examples to highlight how this structure is captured.

hensively represents the pathologies observed in the post-mortem reports. This ontology facilitates accurate classification and analysis of the reported conditions and identifies semantic connections and similarities between various pathologies.

The pathology ontology's development greatly enhances our framework's effectiveness in extracting and interpreting information from the reports. It provides a standardised and systematic approach to categorising and understanding the pathological conditions affecting marine mammals. With this ontology in place, our framework can accurately identify and analyse the different pathologies recorded in the post-mortem reports, contributing to a deeper understanding of the factors influencing the mortality of these animals.

Anatomy Ontology

The anatomy ontology was developed by examining the anatomical terms used in the reports to convey observations. At the highest level of the ontology's "anatomical region" tree, classes represent different organ systems within the body or anatomical regions that are semantically related. For instance, the "integument region" refers to the skin, while the "external region" primarily encompasses external observations beyond the scope of the integument class.

Subclasses at the next level generally represent different types of these regions. A distinction is made between a parent-child relationship and using the "isPartOf" attribute. For example, it would not be accurate to classify "the left valve of the heart" as a subclass of "heart." However, it is still important to capture the relationship between these two regions.

To represent such relationships accurately, the "isPartOf" object property is utilised, which is transitive and asymmetric. This property allows for a more precise representation of instances like the duodenum, which is part of the small intestine, and the small intestine, in turn, is part of the intestines. Therefore, we can deduce that the duodenum is also part of the intestines. This hierarchical structure and using the "isPartOf" attribute ensure the ontology accurately captures the relationships between anatomical regions.

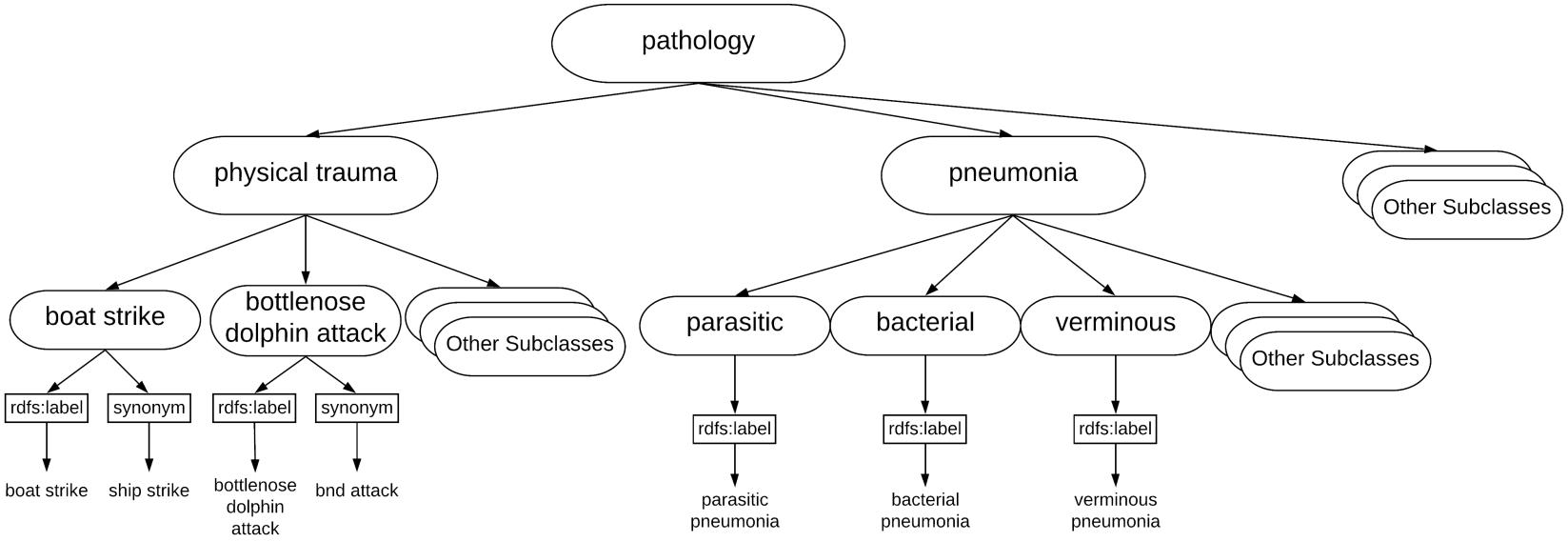


Figure 4.5: Structure of the pathology ontology demonstrated using the physical trauma and pneumonia class examples to highlight how this structure is captured.

Figure 4.6 illustrates the anatomy ontology's structure, showcasing some example anatomies within the alimentary system. The ontology was manually populated and structured based on the headings used in the sections of the gross pathology reports, as shown in Figure 4.2.

By employing the anatomy ontology, our framework gains a deeper understanding of the anatomical context in the post-mortem reports. It allows for a more accurate and structured representation of the observed anatomical regions, facilitating precise retrieval and analysis of the data. The ontology provides a standardised framework for organising and interpreting anatomical observations, contributing to the overall effectiveness of our information retrieval system.

4.2.4 Information Retrieval

Information retrieval consists of several individual components within the framework, including *(a)* lexicon, *(b)* entity-relationship extraction engine, *(c)* anatomy, observation and presence recognition and *(d)* formatting to extract anatomical features, observations and pathologies. This pipeline is shown in the flowchart in Figure 4.1 as well as Algorithm 1.

Lexicons

To facilitate the identification of entities within the post-mortem reports, we generate two lexicons of key terms based on our anatomy, pathology, and observation ontologies, as described in Section 4.2.3. These lexicons play a crucial role in information retrieval by providing a comprehensive list of relevant terms.

The first lexicon, the observation lexicon, is created by parsing the observation and pathology ontologies. We extract the "rdfs:label" and "synonym" attributes from these ontologies' XML files. The observation lexicon captures terms related to observations made in the reports, including both general observations and specific pathological conditions. This is because pathological terms can be used to represent the overall pathology of the specimen and the condition of specific anatomical regions. For instance, if a report mentions a case of physical trauma caused by entanglement, the term "en-

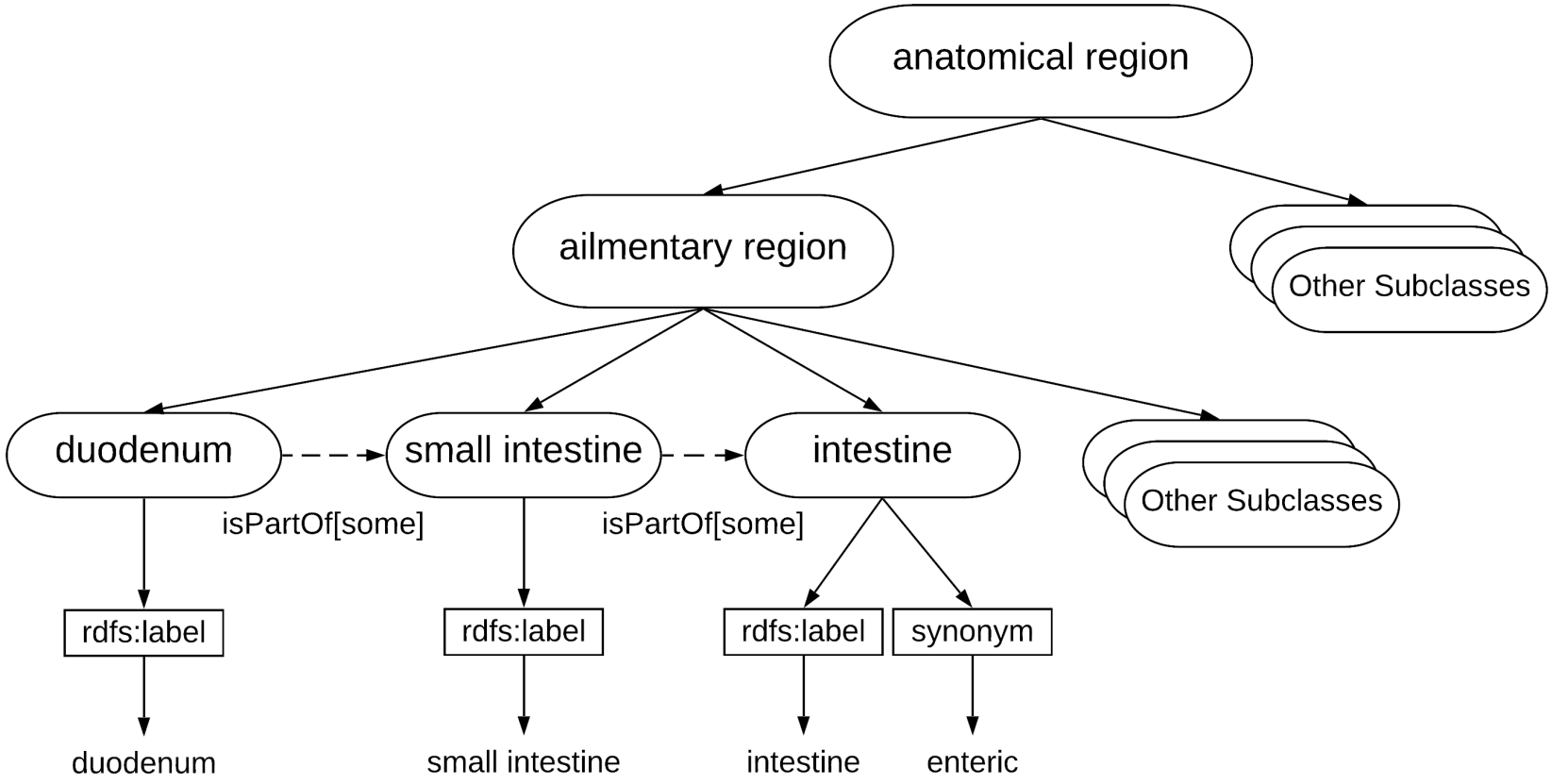


Figure 4.6: Hierarchical structure pf anatomy ontology demonstrated using example ailmentary region subclasses to highlight how this structure is captured.

tangled" can be treated as both an observation and a pathology based on its representation within our ontology.

The second lexicon, the anatomy lexicon, is created by parsing the anatomy ontology. Again, we extract the “rdfs:label” and “synonym” attributes to populate the lexicon. The anatomy lexicon focuses specifically on terms related to anatomical regions within the reports. It enables us to accurately identify and analyse observations pertaining to different organ systems and anatomical regions in a standardised manner.

By incorporating pathological and anatomical terms in the observation lexicon, we ensure that our framework can capture and analyse a wide range of observations, providing a comprehensive understanding of the post-mortem reports. These lexicons serve as valuable resources for entity identification, enabling effective retrieval of relevant information from the reports and supporting subsequent analysis and statistical processing.

Entity-Relationship Extraction Engine

The processing of the reports begins with segmenting the text at the sentence level. Figure 4.2 demonstrates that sections in the reports may not always be clearly delineated by a full stop. Therefore, we also consider new line characters (“\n”) as potential sentence delimiters. After sentence segmentation, we perform word-level tokenisation to break the text into individual words and punctuation marks. We utilise NLTK’s POS Tagger library [123] to assign part-of-speech tags to these tokens.

Next, we apply “Noun Phrase Chunking” (NP-chunking) to group words into meaningful noun phrases. For this purpose, we have developed a feature-based grammar consisting of tag patterns representing entities and entity relationships. This grammar is passed to NLTK’s *Regexparser* library to create chunks corresponding to entities and entity relationships. The regular expression-based grammar we have defined for this task is provided below:

```
NP: {<DT>?<JJ>*<VB.*>*<JJ>*<NN.*>+}
NP: {<NP><CC><NP>}
NP: {<VBD|VBN>}
NP: {<CD><RB>}
```

```
NP: {<NP><NP>}
IN: {<IN>}
REL:{<NP><IN><NP>}
REL:{<NP><TO><NP>}
REL:{<NP><:><NP>}
```

Entity chunks are grouped as noun phrases (NP). The first rule captures cases where there is at least one noun preceded by adjectives (JJ) or verbs (VB), and optionally, a determiner (DT) such as “the” or “a”. Additionally, past tense or past participle verbs used alone are also considered as noun phrases. This ensures that phrases like “right eye: scavenged” are correctly chunked. Furthermore, noun phrases can be linked into a single noun phrase when separated by coordinating conjunction terms such as “and”.

Noun phrases (NPs) are then linked into relationship (REL) chunks based on certain conditions. The first relationship is established when two NPs are adjacent since proximal entities are likely to be related. Prepositions (*e.g.*, “in”) are particularly interesting as they indicate a relationship between the preceding and following entities. The word “to” is also considered a good link between NPs, as phrases like “damage to the left flank” are common. Lastly, we use the colon to capture cases where an anatomical entity is stated followed by observations. Figure 4.2: “Blubber: NAD” shows an example of this.

This grammar is designed to capture relatively simple expressions found in the reports, but it can be expanded to incorporate more complex entity relationships. By applying this grammar, we can effectively identify and extract entities and their relationships from the text, enabling further analysis and processing of the information contained in the reports.

Anatomy, Observation and Presence Recognition

In the processing pipeline, each sentence in a report is examined, and each relationship (REL) chunk is parsed and compared against the anatomical and observational lexicons. If a noun phrase (NP) chunk contains a substring that appears in either lexicon, it is classified as an anatomical entity or an observation, depending on the matching lexicon.

To maintain accuracy in identifying negated terms, the implementation deliberately includes only NP-NP relationships, as defined in our grammar. This approach reduces the occurrence of falsely negated terms. Consequently, a relatively straightforward process can be employed to identify negated words with high accuracy.

For marking negations, we utilise the NLTK mark_negated package. This package appends the suffix "_NEG" to any word between a negation term and certain punctuation marks. A negated statement version is generated for each REL chunk containing an identified observation. The NP chunk that represents the free text observation is then compared to the same chunk after negated terms are marked. If an observational term is negated, it is considered to be absent. For example, in the phrase "no obvious rake marks on the flank," the term "rake marks" would be identified as an observational entity. However, when compared to the negation-marked version of the text ("there are no obvious_NEG rake_NEG marks_NEG on_NEG flank_NEG"), the negation of the observation becomes apparent. In this case, "rake marks" would be classified as "absent." The advantage of this approach is that negated terms are marked only at the relationship level, avoiding incorrect classification of absence due to unrelated negated terms at the sentence level.

When an observational term is recognised but not attributed to an anatomical entity, it is still recorded as either present or absent without being associated with a specific anatomical region. Several reasons may explain why an anatomical entity might not be identified: the term used is not included in the anatomical ontology, the observation is unrelated to an anatomical entity, or the chunking grammar fails to capture relevant NP chunks within a relationship.

By employing these strategies, the system can accurately identify entities, handle negations, and record observations as present or absent, even in cases where a direct link to an anatomical region is not established.

Formatting Findings

The extracted information is consolidated and organised in a Python dictionary, enabling its utilisation in analysis or classification systems. In the dic-

tionary, the anatomical and observation terms are represented as strings, while the presence or absence of an observation is stored as a Boolean value. Here are a few examples illustrating the structure of the dictionary:

```
{
    'anatomy': 'right pectoral fin',
    'observation': 'scavenger damage',
    'presence': True
}
{
    'anatomy': 'epidermis',
    'observation': 'rake marks',
    'presence': False
}
{
    'anatomy': 'skull',
    'observation': 'nad',
    'presence': True
}
```

In this dictionary, the "anatomy" key holds a list of anatomical terms extracted from the reports, such as organs, body regions, or specific anatomical features. The "observation" key contains a list of observation terms extracted from the reports, representing various findings or conditions. The "observations" key further organises the information by mapping each observation term to its corresponding presence (True) or absence (False) value.

This dictionary structure provides a convenient format for further analysis, classification, or processing of the extracted information within Python-based systems.

4.3 Use Case: Bottlenose Dolphin Attacks on Harbour Porpoises

To assess the effectiveness of our approach, we conducted an analysis that focused on observations capable of either confirming or negating a specific

Algorithm 2: Deterministic BDA classification process based on presence or absence of observations.

```
Result: prediction
if Any observation is a BDA term then
    if observation present then
        prediction <- “BDA”;
    else
        prediction <- “Non-BDA”;
    end
    return prediction;
else
    if Any present observation is a GSA or claw mark term then
        prediction <- “Non-BDA”;
        return prediction;
    else
        if Any present observation is a rake mark then
            prediction <- “BDA”;
            return prediction;
        end
        prediction <- “Non-BDA”;
        return prediction;
    end
end
```

pathological finding. We employed the presence or absence of these observations as a basis for classification. In our evaluation, we concentrated on cases involving Bottlenose Dolphin Attacks (BDA) on harbour porpoises, a prevalent cause of death in our dataset. Out of the 193 cases in the SMASS database, 50 listed BDA as the primary finding.

We created a deterministic classifier using the extracted empirical observations. The classifier's classes were established based on explicit mentions of BDA or strong indicators such as "rake marks." In cases where BDA was explicitly mentioned, its presence or absence alone was sufficient to classify the case as either “Non-BDA” or “BDA.” However, the observation “rake marks” could also indicate Grey Seal Attacks (GSA). Therefore, we filtered all observations related to seal attacks and claw marks (an indicator of GSA). We assumed that if both a GSA and a BDA were present, there should be an explicit mention of BDA. If a document did not adhere to these classification rules, the presence of "rake marks" would result in a “BDA” classification.

Metrics	Cumulative	BDA	Non-BDA
Accuracy	0.83	-	-
ROC-AUC	0.77	-	-
Recall	-	0.64	0.90
Precision	-	0.70	0.88
F1-score	-	0.67	0.89
Weighted F1-score	0.83	-	-
Support	193	50	143

Table 4.1: BDA classifier performance evaluation metric scores.

The sequential decision-making process of the deterministic classifier is depicted in Algorithm 2.

To evaluate the classifier's performance, we used the cause of death information stored in the SMASS database as the ground truth. By comparing the classifier's classifications with the known causes of death, we could assess the accuracy and effectiveness of our approach.

The results of the BDA use case are presented in Table 4.1. Our approach achieved an overall accuracy of 83.4% and an F1-score of 0.83. Specifically, for BDA classification, we achieved a precision of 0.70, recall of 0.64, and an F1-score of 0.67. For Non-BDA classification, the precision was0.88, recall was 0.90, and the F1-score was 0.89. Out of the 193 reports analysed, 50 were classified as BDA cases, while 143 were classified as Non-BDA cases based on the cause of death information from the SMASS database.

To evaluate the classifier's performance, we generated a Receiver Operating Characteristic (ROC) curve Figure 4.7 using the precision and recall values for BDA classification mentioned above. The area under the curve (AUC) was calculated to be 0.77, indicating a reasonable level of performance. However, it is important to note that the labelling based on cause of death has limitations, as there are instances where a BDA occurred, but another finding was identified as the cause of death. This can lead to false positive (FP) classifications, as BDA terms and indicators may still be mentioned in such cases. The confusion matrix in Figure 4.8 visually represents the classification results.

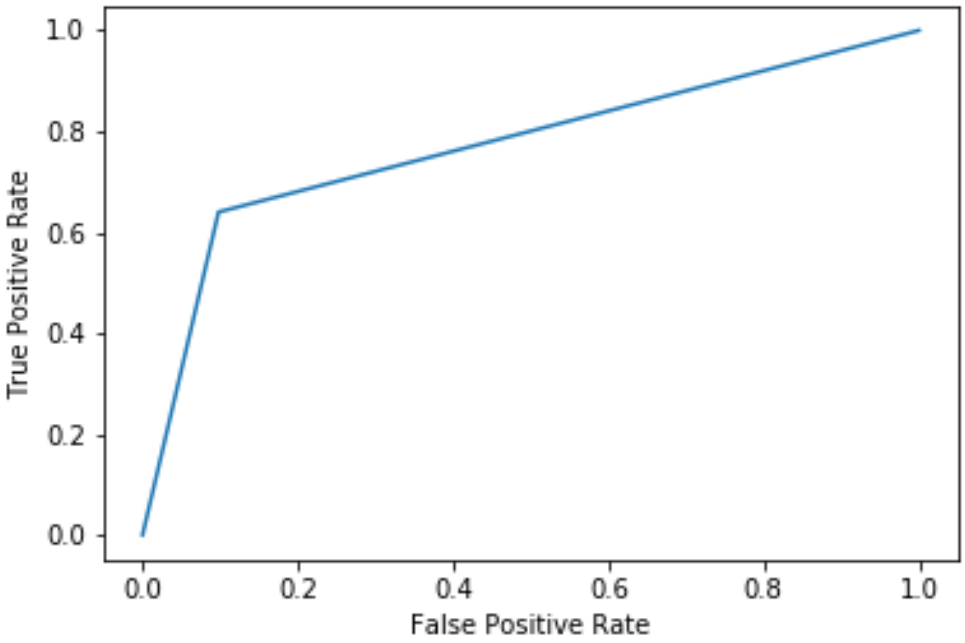


Figure 4.7: ROC-AUC curve of BDA classifier predictions. AUC = 0.771

Due to the deterministic nature of our classifier, there are three possible causes for incorrect classifications. First, there may be a separate and more significant finding that caused death, even though a BDA occurred. Second, a significant term may be used outside of its intended scope. Lastly, a significant finding may not be successfully identified by the entity-relationship engine, leading to missed classifications. These factors contribute to the challenges of accurately classifying cases using our approach.

During the analysis of false positives (FPs), we observed incorrect detections in cases where statements like “rake marks, assumed bird” were present. This highlights the challenge of dealing with varying terminology usage among pathologists. Some FPs were also attributed to instances where BDA rake marks were described as "healed" or “healing.” On the other hand, false negatives (FNs) occurred when there were no explicit mentions of BDAs, despite statements such as “no obvious rake marks.” This suggests that other indicators of BDA may present even when rake marks are absent.

The relatively low recall (0.64) and precision (0.70) scores in BDA classifications can be attributed to the simplicity of the feature-based grammar used and the underestimation of the total number of true positives (TPs).

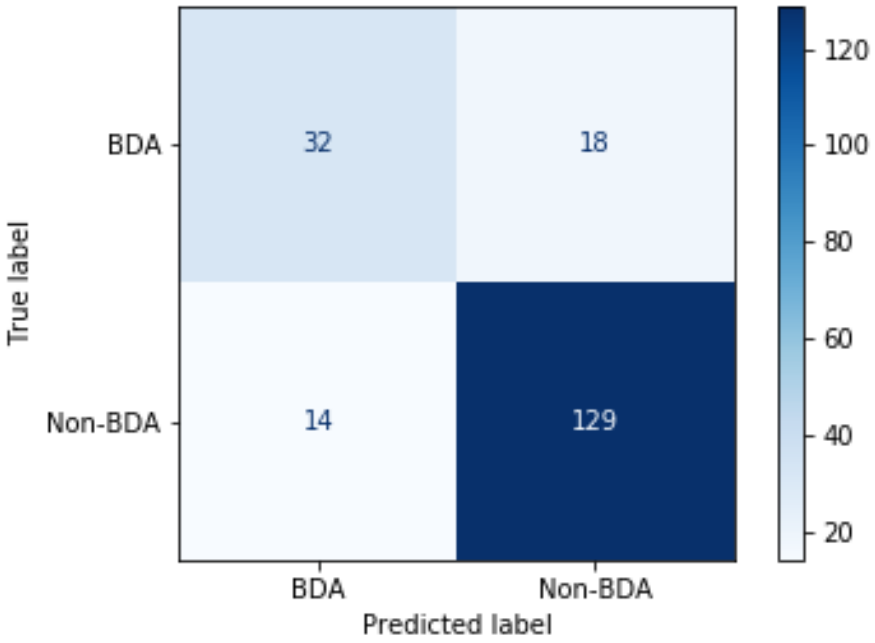


Figure 4.8: Confusion matrix of results from our deterministic BDA classifier predictions.

Analysis of FPs and FNs also revealed cases where the feature-based grammar in the entity-relationship engine did not capture significant phrases. This indicates the need for further refinement and expansion of the grammar to improve the classifier's performance.

In contrast, the precision and recall scores for Non-BDAs (0.88 and 0.90, respectively) were considerably higher. This was mainly due to excluding Grey Seal Attacks (GSAs) in the deterministic classifier. An inclusion/exclusion-based determiner effectively increased confidence in positive classifications, making the insights obtained more robust. Despite minor shortcomings in the grammar used, the results are highly promising for future work.

It is worth considering that a necropsy report contains additional fields related to morphology, diagnostic confidence, and other free-text sections like histopathology reports and conclusion sections. We can define more accurate, complex, and inclusive determiners by incorporating these relevant fields and applying a similar information retrieval process to other sections. This would result in higher confidence in positive or negative classifications.

While using cause of death as a label led to lower performance metrics than anticipated, it is important to note that without a manually labelled dataset, we can still better understand the classifier's characteristics. Future work should explore incorporating additional report fields and refine the methodology to improve the classifier's accuracy and comprehensiveness.

4.4 Conclusions

This chapter presents our ontology-based text mining framework for analysing marine mammal gross pathology reports, demonstrating its efficacy in a case study involving BDAs on harbour porpoises. While our approach performed well, there are weaknesses in problem formulation and the entity-relationship engine, occasionally leading to incorrect classifications.

It's worth noting that while we have applied this approach to marine mammal gross pathology reports, the agnostic way in which we develop the ontologies allows for the potential extension of similar techniques to reports of different types, including histology and bacteriology reports, and even reports on different species altogether. Further discussions on these findings and their implications can be found in Chapter 9.

5 Imaging Datasets

In this project, we have developed a range of datasets for various experiments. This chapter provides a detailed description of each dataset and its unique characteristics.

5.1 Salmon Gill Pox Virus Dataset

This dataset was acquired from a separate PhD project studying poxvirus in Atlantic Salmon gills funded by Cooke Aquaculture and the University of Stirling, with the former providing the samples. The dataset consisted of images of freshwater Atlantic salmon, pre-smolt. It allowed us to begin the preliminary exploratory phase of this project. Comprising a total of five Whole Slide Images (WSIs), it provided sufficient diversity in the severity of pathology to develop machine-learning methods for our hyperplasia analysis. Additionally, it was used to evaluate which techniques could best accentuate the features necessary for accurate hyperplasia analysis.

Each WSI was divided into tiles of size 1024x1024 pixels before being programmatically evaluated as relevant or irrelevant. The first step involved checking if a region contained any information, given that many WSIs contain empty regions. Once the tile was converted to greyscale, this was done by checking if any pixel in the image was not equal to 0. Tiles were blurred, and a threshold was applied to segment tissue regions. Tiles that contained at least 20% tissue were deemed to contain enough information for hyperplasia analysis and included in the dataset. This process is shown in Figure 5.2.

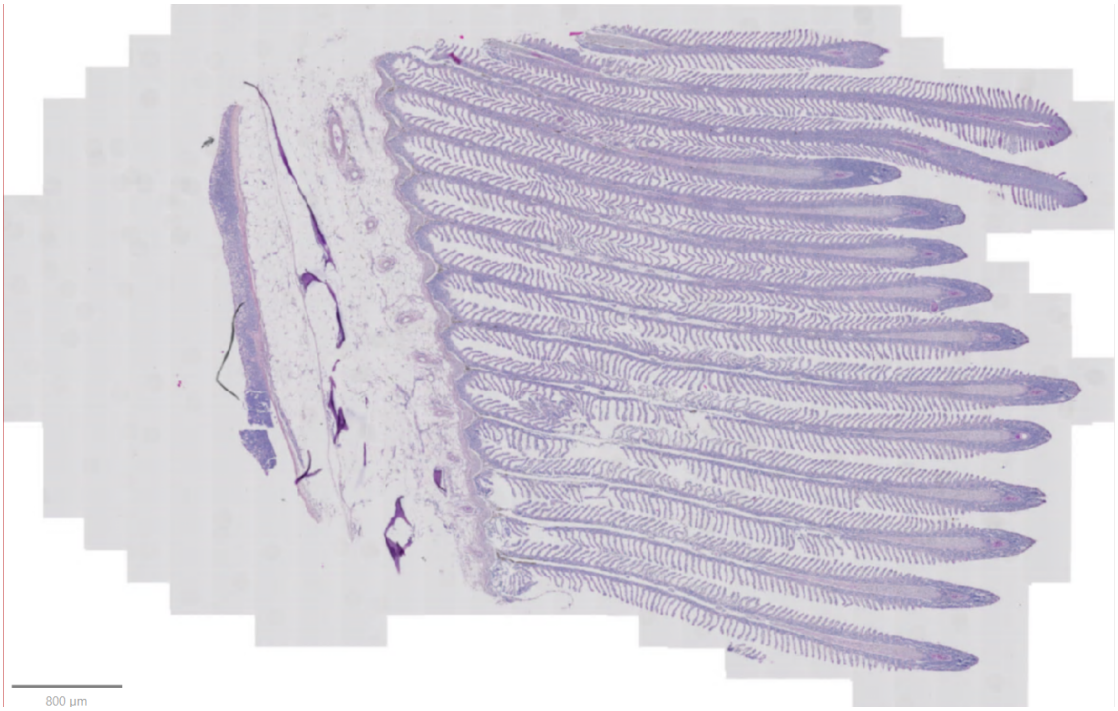


Figure 5.1: Example WSI from the Salmon Gill Pox Virus Dataset. This small dataset included images representing a diverse range of severity of epithelial hyperplasia. This example in this figure shows a relatively healthy H&E stained gill image.

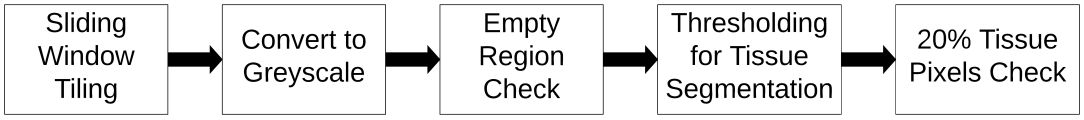


Figure 5.2: The processing pipeline for creating our tiled datasets.

Tiles were then manually labelled before being verified by an expert pathologist. Labelling was done based on the severity of hyperplasia exhibited in a tile and classified as normal, mild, moderate, or severe. Our paper used this dataset, which evaluated deep learning and signal processing-based methods for hyperplasia classification [4]. The final version of the dataset consisted of 1465 tiles. The moderate class was slightly underrepresented, with only 169 samples.

Category	Train Counts	Val Counts
Normal	272	124
Mild	319	139
Moderate	122	47
Severe	312	130
Total	**1025**	**440**

Table 5.1: Train and Validation Counts for Salmon Gill Pox Virus Dataset

5.2 Expanded Evaluation Dataset

An extended version of the dataset was developed using a further 13 WSIs provided by the archive of the Institute of Aquaculture at the University of Stirling. The same pipeline was used to include and exclude tiles before labelling. This dataset contained a much wider variety of pathologies and additional signals such as varying Hematoxylin and Eosin staining intensities. As well as the labelled version of the dataset, which was used for the classification of hyperplasia, an unlabelled version was also developed to develop the anomaly detection methods described later in the thesis. The number of tiles associated with each class and the split between training and testing sets can be seen in Table 5.2.

Figure 5.3: Example WSI from the Expanded Evaluation Dataset. The expanded evaluation set included several images from the internal archive of the Institute of Aquaculture at the University of Stirling. This specific example shows more severe pathology, and has much darker staining when compared to the image in Figure 5.1.

Category	Train Counts	Val Counts
Normal	2530	1024
Mild	5425	1442
Moderate	625	208
Severe	501	460
Total	**9081**	**3134**

Table 5.2: Train and Validation Counts for Expanded Evaluation Dataset

5.3 Gill Health Project Dataset

Pharmaq Analytiq is an Inverness-based veterinary services and diagnostic technology company. The company (previously Fish Vet Group) provides aquaculture services in the United Kingdom and Ireland. One of the services offered is the histopathological examination of tissue samples taken from fish. Over time, they have amassed a sizable collection of histology Whole Slide Images (WSIs), complete with free-text reports and, in some cases, lesion scores for each image in the collection. WSIs were collected as a result of their inclusion in the SRUC's "Gill Health Project". The samples used in this work were from saltwater salmon which were sampled from sites periodically.

This dataset comprises 20 WSIs obtained from the SRUC's Gill Health Project [124]. Slides had been manually scored for a variety of lesions, including hyperplasia. This whole slide score allowed us to evaluate and compare the scores produced by pathologists and those produced by various configurations of our framework. The number of slides associated with each severity score is represented in Figure 5.5.

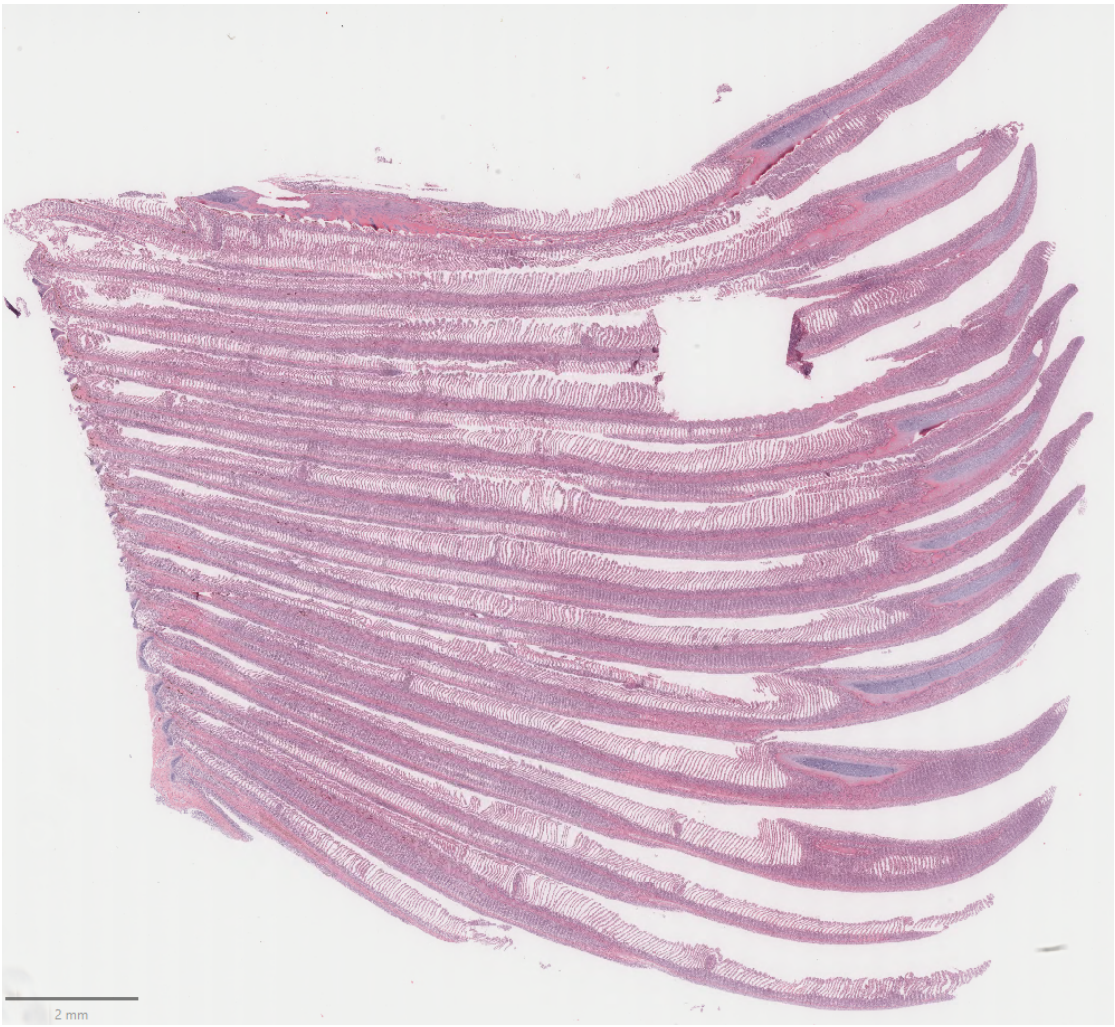


Figure 5.4: An example slide from the Gill Health Project dataset. This example exhibits relatively little pathology, and the staining is paler than the examples shown in Figures 5.1 and 5.3. This example shows areas of focal hyperplasia.

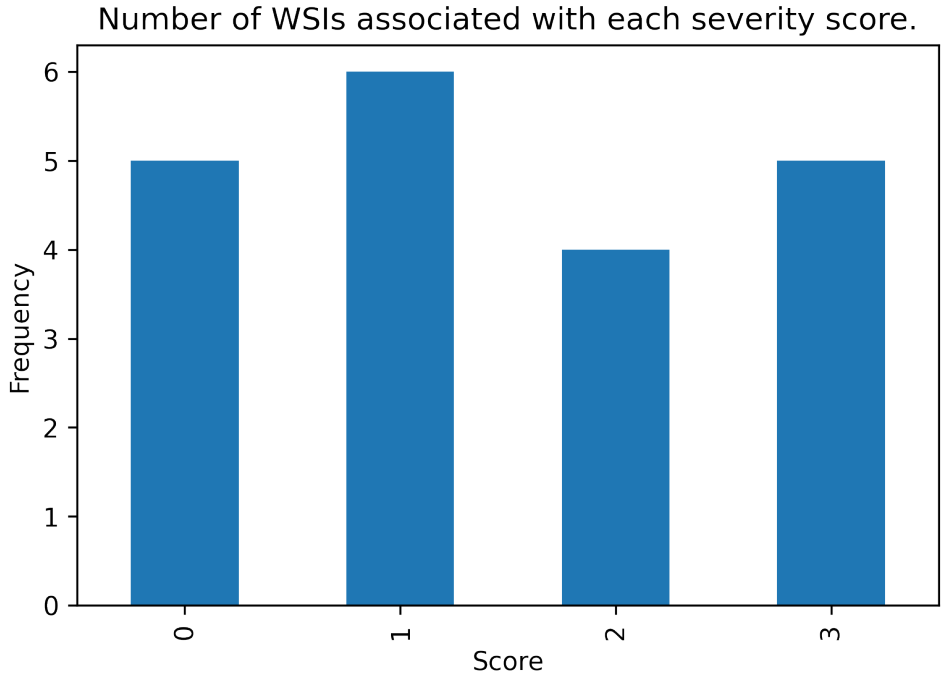


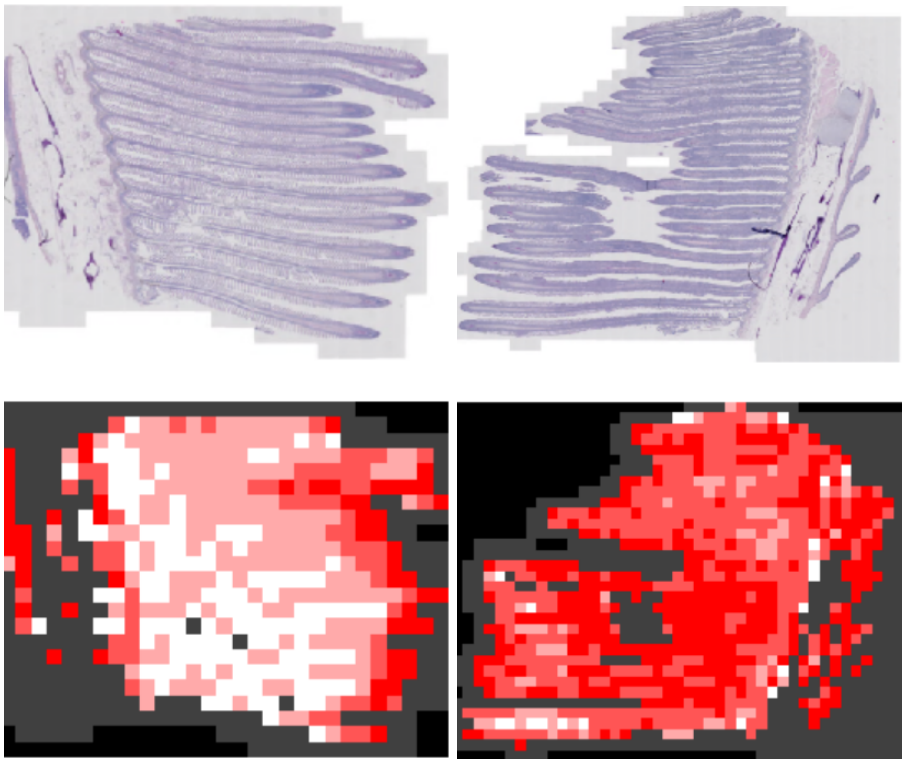
Figure 5.5: The number of tiles associated with each level of severity in the Gill Health Project dataset. 0 is healthy, 1 is mild hyperplasia, 2 is moderate, and 3 is severe hyperplasia.

6 Empirical Wavelet Transforms for Hyperplasia Score Classification

Tissue samples obtained from fish during post-mortem provide valuable insights into the animal's condition at the time of death. Histopathology plays a critical role in this analysis as it involves examining tissue changes at the microscopic level. One particular lesion of interest is epithelial hyperplasia, which signifies excessive cell proliferation and strongly indicates local irritation, infectious disease, or poor water quality. This chapter focuses on classifying the severity of hyperplasia in Atlantic Salmon whole-slide image (WSI) tiles. We propose a novel classification approach incorporating pathology's domain expertise into the image processing pipeline to achieve this, given that Automating the classification of hyperplasia in WSIs can be achieved through frequency domain texture analysis. We introduce a unique feature engineering technique based on Empirical Wavelet Transform (EWT), which involves analysing the subband statistics of the transformed images. The feature representation obtained from EWT and a fully connected neural network (FCNN) form the basis of our classification methodology.

A key aspect of our approach is the ability to identify the locally affected regions within a WSI. We accomplish this using a tile-based method that allows us to visualise and understand the specific regions contributing to the overall classification. By extracting fine-grained information, our methodology enables a more comprehensive analysis and interpretation of the results. Figure 6.1 illustrates an example output of our proposed algorithm, showcasing its effectiveness in identifying and characterising hyperplasia in WSIs.

Hyperplasia in the gills significantly impacts the microscopic anatomy of



(a) Normal/Mild (b) Severe

Figure 6.1: A heatmap of our approach applied to gill WSIs. White indicates healthy and red indicates severe hyperplasia. Grey areas indicate regions which have been elimated in the thresholding steps of preprocessing.

the tissue. Various factors, such as local irritation, infectious disease, or poor water quality can trigger epithelial hyperplasia. The tissue forms a comb-like structure in healthy gills, with primary lamellae resembling teeth on a comb. These primary lamellae further branch into secondary lamellae, as depicted in Figures 6.1 & 6.2. A healthy gill's secondary lamellae consist of a linear channel filled with red blood cells, surrounded by a thin layer of cells known as the epithelium. In mild to moderate hyperplasia cases, the number of cells in the epithelial layer increases, resulting in a reduction of space between the secondary lamellae. The secondary lamellae fuse completely in severe cases, as illustrated in Figure 6.2. Severe hyperplasia severely impairs the gill's respiratory function and its ability to extract oxygen from water. Therefore, accurately measuring and assessing hyperplasia in gills is crucial for understanding the overall health of the fish.

Automating the classification of hyperplasia in WSIs can be achieved through frequency domain texture analysis. However, traditional image decomposi-

tion methods have limitations. Fixed template-based approaches, such as wavelets, lack flexibility due to their rigid template structures. On the other hand, data-driven approaches like Empirical Mode Decomposition (EMD) [96] lack a solid mathematical foundation. Gilles et al. address the lack of mathematical foundation, proposing the empirical wavelet transform as an adaptive signal decomposition method [40, 97].

In this work, we have used the LP transform to decompose an image into its different frequency components, capturing both local and global variations in the image data. Unlike conventional wavelet-based approaches, the empirical 2D LP transform introduces adaptability in its filtering process, allowing it to capture image features across a wide range of scales effectively.

Extraction methods such as CNNs and our approach enable us to create a tile-based classification framework for the generation of statistics. Our approach quantifies the severity of various lesions within an individual whole-slide image (WSI), thereby creating a model of the gill's condition. The fine-grained understanding of a gill's health allows us to explore more sophisticated methods of scoring lesions. Such a model can serve as a valuable complement to other pathological or epidemiological data. As such, the following contributions are made in this chapter:

- We introduce a unique data-driven approach for measuring hyperplasia in gill histology images, filling a gap in the existing literature. Our approach not only provides a means to quantify hyperplasia but also enables the generation of visualisations to aid in the understanding and interpretation of the results. Our approach also allows us to differentiate between cases of focal and diffuse hyperplasia at the WSI level.
- We propose a novel parametric feature generation method, combined with a fully connected neural network (FCNN), that translates expert knowledge into mathematically explainable features. This method utilises both non-adaptive wavelets and Empirical Wavelet Transform (EWT), capturing relevant information from the images. By incorporating expert knowledge into the feature engineering process, we enhance the interpretability and explainability of the generated features.
- Our methodology introduces a scale and rotation-agnostic representation of image features, enabling the development of effective mod-

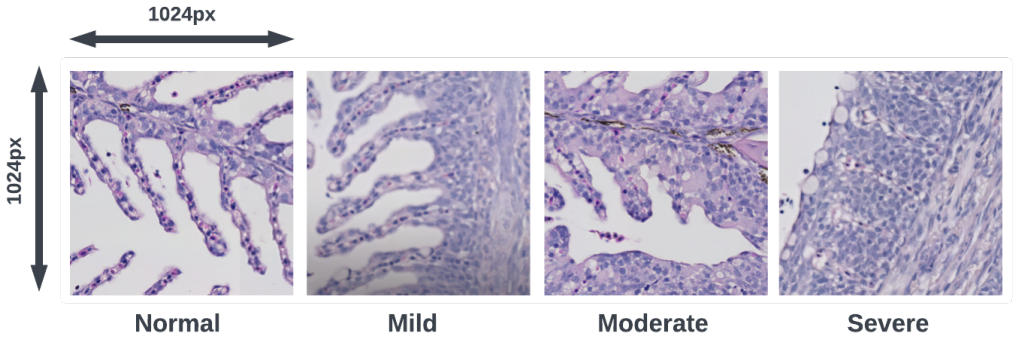


Figure 6.2: Tiles extracted from gill WSIs. From left to right - Normal, Mild, Moderate, and Severe cases of hyperplasia. The epithelial layer surrounding the red cells is very thin in the Normal case, but thicker in the Mild and Moderate tiles. The secondary lamellae are indistinguishable in the severe case, except for the lines of red cells which are still visible.

els with small datasets. This approach overcomes limitations imposed by variations in image scale and rotation, ensuring robust and reliable analysis results.

By producing our method for lesion recognition in gill histology images, we contribute to the understanding and characterisation of gill health. Our data-driven approach, coupled with novel feature generation techniques and scale/rotation invariance, opens avenues for improved analysis, interpretation, and modelling in the field. These contributions advance the state-of-the-art in fish gill pathology and provide valuable tools for researchers and practitioners in the domain.

6.1 Methodology

The core of our pipeline consists of three main sections: preprocessing, subband image generation using 2D LP-EWT, and parametric feature calculation, as illustrated in Figure 6.3. These steps form the foundation of our approach to incorporate expert knowledge as features in the machine learning models.

To ensure the accuracy and reliability of our classification, we collaborated closely with an expert pathologist to establish the ground truth for

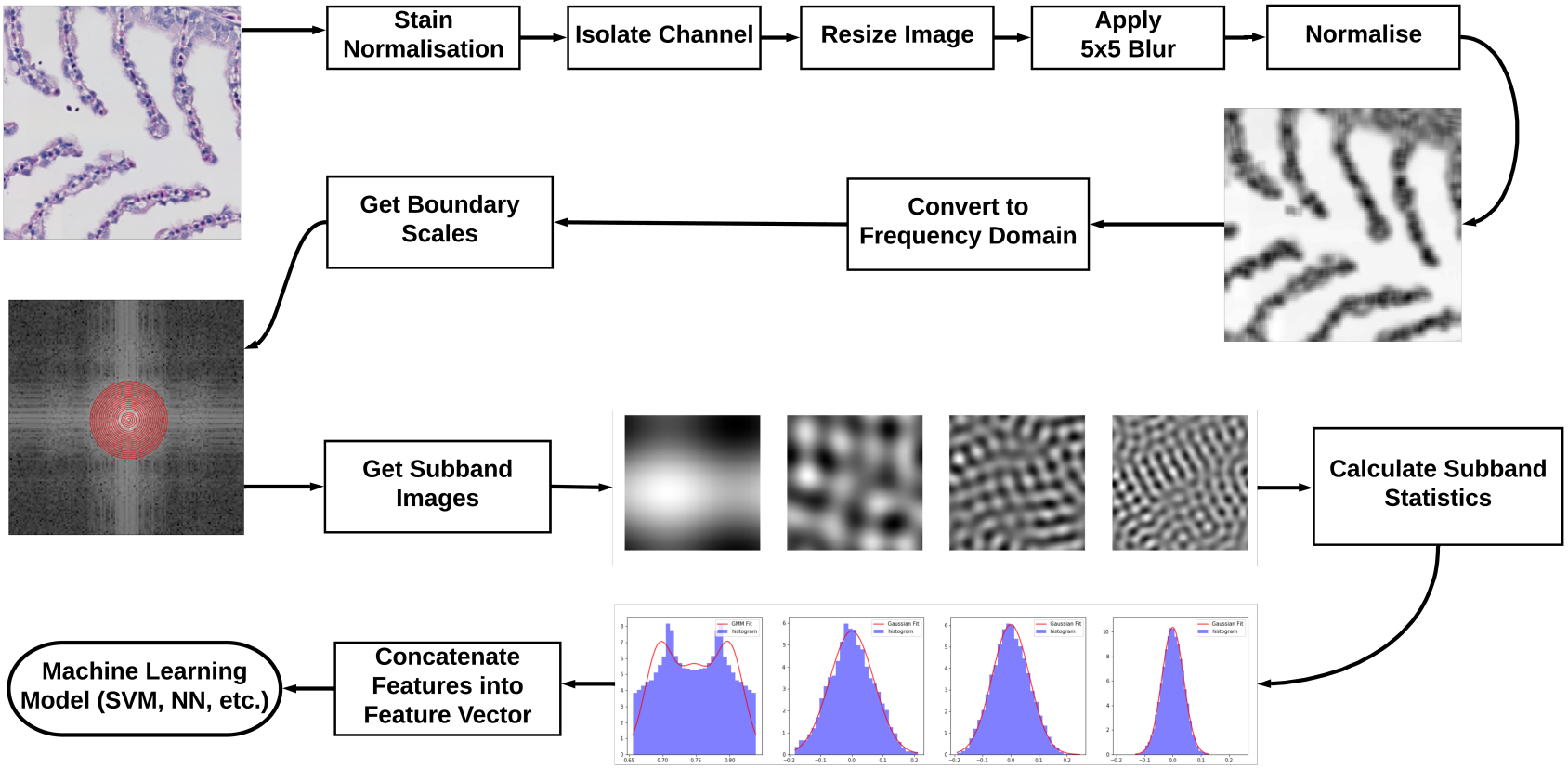


Figure 6.3: A flowchart showing our approach: preprocessing, EWT, and parametric feature generation. A 1024x1024 tile goes through the preprocessing steps outlined in the flowchart, before extracting features from subbands produced using an empirical wavelet transform. A Gaussian Mixture Model fit to the lowest subband is shown as well as Gaussian distributions for 3 other subbands.

hyperplasia. Through consultations, the expert pathologist identified several components of gill histology slides that serve as differentiators between healthy/normal, mild, moderate, and severe cases of hyperplasia. Indicators of hyperplasia included an increase in the overall tissue area, a decrease in the space between secondary lamellae, and a shift in tissue colour.

To select the most relevant features that effectively represent these characteristics, we evaluated potential features and made inclusion or exclusion decisions based on their ability to capture the desired attributes.

The first step in the pipeline, preprocessing, is concerned with accentuating the morphological characteristics of the image while normalising non-anatomical or pathological elements which may vary from slide to slide, such as stain intensity. The EWT is then applied to the separate RGB channels of the tiles, allowing us to capture the image's colour and overall texture information. We then use our parametric feature engineering technique to extract subband statistics from the resultant subband images.

6.1.1 Preprocessing

We resized each image from its original size of 1024×1024 to a reduced size of 128×128. These dimensions were chosen due to the the massive reduction in computational complexity, while still effectively representing the morphological features of the image. This resizing was achieved using pixel area relation resampling, preserving the key details while reducing computational complexity. Subsequently, we applied a low-pass 5×5 average blurring filter to smoothen the image and reduce noise. This blurring operation helps to enhance the overall image quality and improve subsequent analysis results. Normalisation was another important step in our preprocessing pipeline. By applying and evaluating a variety of stain normalisation techniques, we standardised the intensity values across the image, ensuring that the data distribution is consistent and facilitating fair comparisons between different tiles.

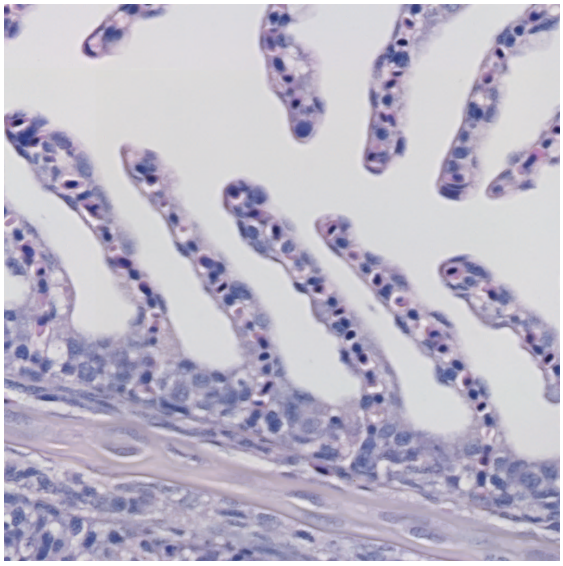


Figure 6.4: The 1024x1024 reference image used in the stain normalisation processes. All tiles have had colours matched to this image using the various stain normalisation methods used in this project.

6.1.2 Stain Normalisation

Histology images stained with Hematoxylin and Eosin (H&E) are extensively utilised in medical and biological research. However, variations in staining protocols and imaging conditions can introduce inconsistencies in the colour appearance of these images.

Several stain normalisation methods have been explored to ensure reliable and consistent analysis. We initially used Vahadane normalisation [4, 125], however, we have since empirically established that other methods are more effective for our purposes. All methods discussed involve the use of a representative image for all other images to be matched to. In this context, we discuss three popular stain normalisation methods used in H&E-stained images:

Macenko et al. [76] assumed that every pixel represents a singular stain vector. The first step involves converting the image to optical density, or OD space, before using singular value decomposition to identify the vector associated with each stain component: hematoxylin and eosin. By estimating stain concentration and colour vectors, Macenko's method effectively normalises the colour appearance of stained images.

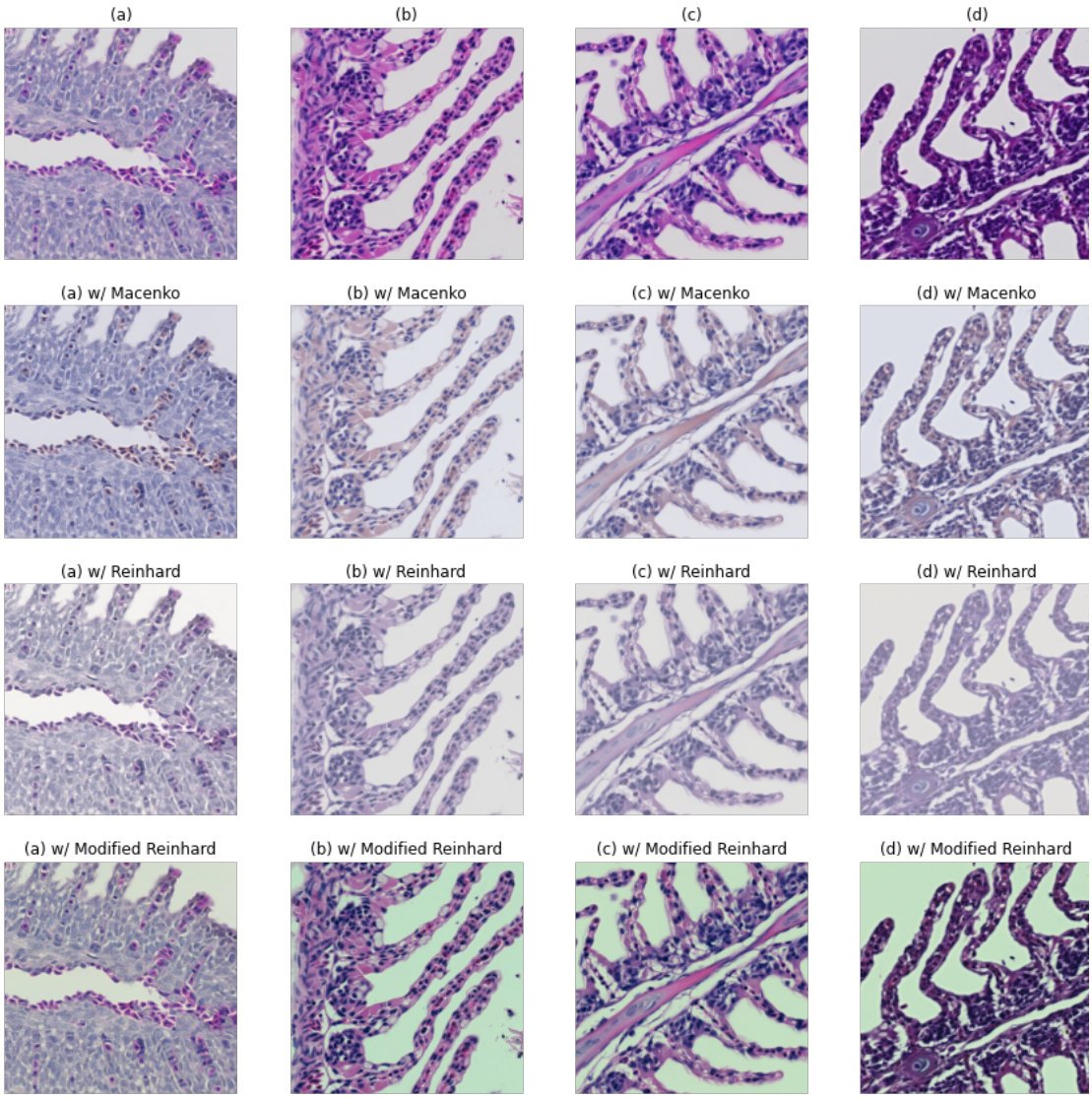


Figure 6.5: Stain normalisation methods demonstrated on three tiles from different slides with varying stain intensities. Tiles are all of size 1024x1024. Reference image in Figure 6.4 used for stain normalisation.

Reinhard et al. [75] introduced an approach to align the colour distribution of an image with a target image using a linear transform in a perceptual colourspace known as the $l\alpha\beta$ colourspace developed by Ruderman et al [126]. The goal is to ensure that the means and standard deviations of each colour channel in both images are matched in that colour space.

A modified version of Reinhard et al.'s algorithm was presented by Roy et al. in 2021 [127]. They aimed to address some of the known shortcomings of Reinhard normalisation, namely that Reinhard normalisation does not

preserve background luminance and that the method sometimes produces a poorly contrasted image where the reference image has less contrast than the target.

All three of these methods were applied using the torchstain python library [2].

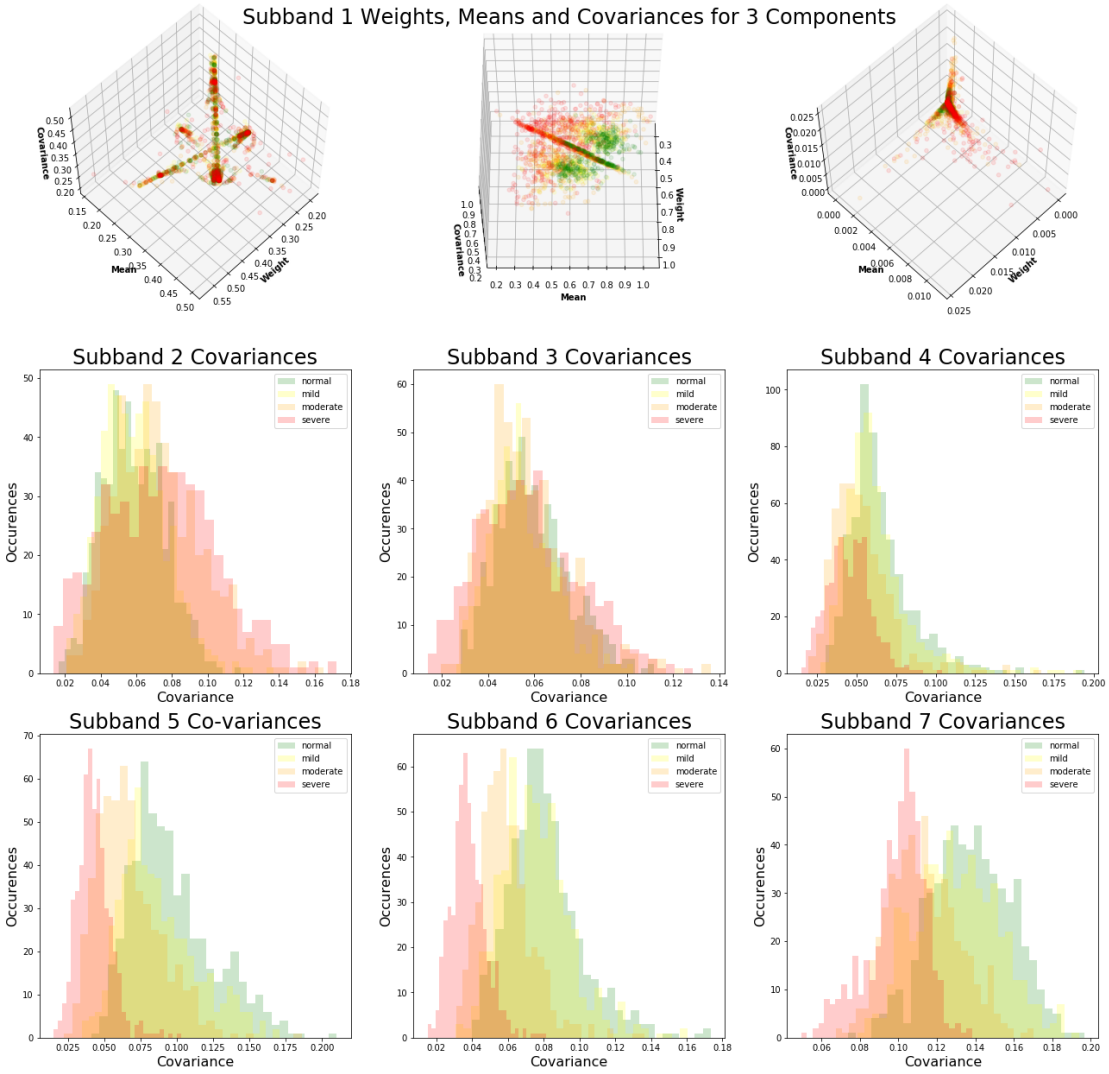


Figure 6.6: Scatter plots of the 3 components of the GMM model fit to the tile image, the three dimensions being weight, mean, and co-variance. The histograms show how covariances produced from a Gaussian model vary between classes.

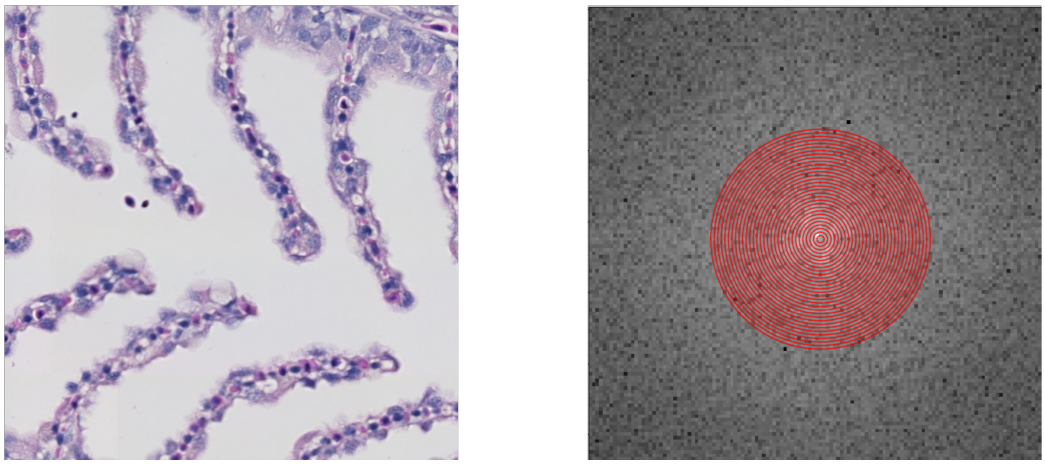


Figure 6.7: A tile taken from a WSI next to its LP EWT. Subband boundaries are shown in red.

6.1.3 Parametric Feature Engineering

Given that we identified colour and texture as relevant characteristics for our approach to capture, the LP-EWT method described above was applied to separate colour channels of each tile. We applied a parametric method to the subband images generated by EWT to produce the features used for model training and testing.

Our approach produces, n, subband images when we apply LP-EWT to the tile. We then extract statistical features from each image by fitting bespoke probability distribution functions (PDF) to flattened representations of the tiles' pixels. A Gaussian Mixture Model (GMM) PDF is fit to the 1st image, with a given number of, $\mathcal{K}$, components. For each k^{th} component, a mean, μ_k, a variance, σ_k, and a weight, ϕ_k are produced. In our approach we have empirically set $\mathcal{K}$ to be 3, meaning for the 1st subband image we obtain 9 features. Equation 6.1 describe the GMM PDF-based feature generation process.

$$p(x) = \sum_{i=1}^{K} \phi_i \mathcal{N}\left(x \mid \mu_i, \sigma_i\right), \tag{6.1}$$

where,

$$\mathcal{N}\left(x \mid \mu_i, \sigma_i\right) = \frac{1}{\sigma_i \sqrt{2\pi}} \exp\left(-\frac{(x-\mu_i)^2}{2\sigma_i^2}\right),$$

such that,

$$\sum_{i=1}^{K} \phi_i = 1.$$

For all subband images except the first, a Gaussian PDF is fit based on the pixel values (shown in Eq. 6.2). The σ is kept and used for the feature vector, while the mean, μ, is discarded as $\mu = 0$ in all cases.

$$p(x; \mu, \sigma) = \frac{1}{\sigma\sqrt{2\pi}} e^{-\frac{1}{2}\left(\frac{x-\mu}{\sigma}\right)^2}. \tag{6.2}$$

This process is applied to three different representations of the tile using the red, green, and blue colour channels. Given that n=20 in this work, we generate 29 features for each colour channel and 87 in total.

The features extracted from the 1st subband image using the GMM PDF are concatenated and organised.

$$\mathcal{M} = (\mu_0, ..., \mu_k, \sigma_0, ..., \sigma_k, \phi_0, ..., \phi_k). \tag{6.3}$$

The features extracted from the other subband images from each channel ($\mathcal{H}_r, \mathcal{H}_g, \mathcal{H}_b$) are then also concatenated to produce the feature vector, $\mathcal{H}$:

$$\mathcal{H} = (\mathcal{H}_r, \mathcal{H}_g, \mathcal{H}_b), \tag{6.4}$$

where,

$$\mathcal{H}_r = (\mathcal{M}_r, \sigma_{r1}, \sigma_{r2}, ...\sigma_{rn}),$$
$$\mathcal{H}_g = (\mathcal{M}_g, \sigma_{g1}, \sigma_{g2}, ...\sigma_{gn}),$$
$$\mathcal{H}_b = (\mathcal{M}_b, \sigma_{b1}, \sigma_{b2}, ...\sigma_{bn}).$$

The feature vector ($\mathcal{H}$) is then passed to a fully connected neural network (FCNN) with 3 hidden layers of 1024 neurons. We analysed the features generated from a 7 subband image EWT on a balanced subset of our Salmon Gill Pox Virus Dataset dataset. The visualisation in Figure 6.6 shows scatter plots of the components generated from parameters generated by fitting a 3-component GMM on the lowest subband image, and histograms generated from the co-variants of the other subband images. This visualisation demonstrates a separation between classes that, when combined, can be

leveraged for the purposes of classification.

6.2 Results

All experiments in this section were run on an AMD X86_64 with an RTX2080Ti. The experiment evaluated the performance of different stain normalisation approaches for the models: InceptionV3, LP-EWT, and ResNet18. Transfer learning was not used. The loss function used was cross-entropy loss. Stochastic Gradient Descent (SGD) was the optimiser used. The results, shown in Table 6.1, revealed that Reinhard stain normalisation was the best approach for all three models. When studying how the models performed when classifying individual severity classes, it becomes quickly apparent that the data imbalance has lead to difficulties in correctly classifying the moderate class, given the poor performance across all models and metrics. It is important to also highlight the limitations of representing a continuous attribute such as hyperplasia as a discrete category. We discuss these findings, their interpretations, and their implications further in Chapter 9. In its current form, our model can effectively produce tile-level information relating to the severity of hyperplasia, however, our approach does not yet discriminate between relevant and non-relevant tissue, which we address in Chapter 7.

6.3 Conclusions

We developed and evaluated a method of feature extraction which used empirical wavelet transforms alongside probability density functions. Our method utilised both Gaussian and Gaussian Mixture models to extract features from subband images in order to train and test our neural network model. We compared our model to several deep learning methods and found that our approach achieved comparable performance. Furthermore, we evaluated three stain normalisation methods across all models evaluated: Macenko, Reinhard, and Modified Reinhard. We found that Reinhard stain normalisation performed best among the methods evaluated. This analysis supplements our previous work research, which showcases that

Model		Macenko			Reinhard			Modified Reinhard		
		Precision	Recall	F1-Score	Precision	Recall	F1-Score	Precision	Recall	F1-Score
EWT-LP	0 (Normal)	0.67	0.22	0.33	0.57	0.41	0.48	0.5	0.4	0.44
	1 (Mild)	0.56	0.92	0.69	0.58	0.78	0.66	0.57	0.66	0.61
	2 (Moderate)	0.45	0.16	0.23	0.44	0.13	0.2	0.32	0.37	0.34
	3 (Severe)	0.95	0.7	0.81	0.89	0.78	0.83	0.85	0.75	0.8
	macro avg	0.66	0.5	0.52	0.62	0.53	0.54	0.56	0.54	0.55
	weighted avg	0.64	0.61	0.56	0.61	0.62	0.6	0.57	0.57	0.56
	accuracy		0.61			0.62			0.57	
ResNet18	0 (Normal)	0.68	0.38	0.49	0.67	0.44	0.53	0.66	0.33	0.45
	1 (Mild)	0.61	0.86	0.72	0.61	0.86	0.71	0.61	0.9	0.73
	2 (Moderate)	0.34	0.27	0.3	0.29	0.21	0.24	0.35	0.35	0.35
	3 (Severe)	0.88	0.68	0.77	0.93	0.57	0.71	0.88	0.52	0.65
	macro avg	0.63	0.55	0.57	0.62	0.52	0.55	0.63	0.53	0.54
	weighted avg	0.65	0.64	0.62	0.65	0.64	0.62	0.65	0.62	0.6
	accuracy		0.64			0.64			0.62	
InceptionV3	0 (Normal)	0.83	0.29	0.43	0.67	0.71	0.69	0.51	0.87	0.64
	1 (Mild)	0.61	0.88	0.72	0.68	0.74	0.71	0.69	0.64	0.67
	2 (Moderate)	0.33	0.57	0.42	0.37	0.23	0.28	0.29	0.03	0.05
	3 (Severe)	0.87	0.68	0.77	0.91	0.71	0.8	0.81	0.03	0.05
	macro avg	0.66	0.61	0.58	0.66	0.6	0.62	0.57	0.39	0.35
	weighted avg	0.7	0.64	0.61	0.69	0.69	0.69	0.62	0.59	0.53
	accuracy		0.64			0.69			0.59	

Table 6.1: Table of performance metrics for different combinations of classification models and stain normalisation techniques.

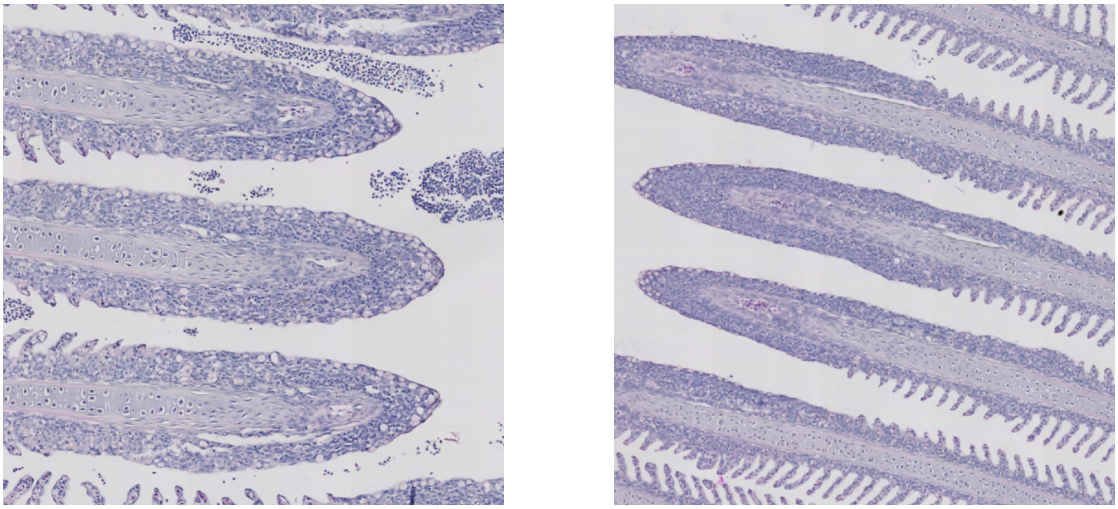
our EWT-LP method not only achieves much faster training processing times but also maintains comparable accuracy scores [4].

7 Context-Aware Hyperplasia Analysis using Variational Autoencoders

The dataset used in the previous chapter consists of tiles of lamellar tissue with varying degrees and types of pathology. This allowed us to train more effectively and test models that can perform lesion-based analysis. However, the sliding window method used in the previous chapter to generate metrics and visualisations lacks a contextual understanding of what is and what is not lamellar tissue. This context-agnostic approach has limitations as it evaluates non-lamellar tissue, which should not be considered when generating a global metric for the level of hyperplasia in a slide. Similarly, this approach's visualisations may highlight irrelevant tissue areas as abnormal. An example of this can be seen in Figure 6.1, where the normal/mild case contains areas identified as severe. When analysed, it was found that these anomalies were either excess tissue from the gill arch, or cells at the tips of the primary lamellae, which resemble severe hyperplasia when considered out of context. Figure 7.1 shows how the tips of primary lamellae lose the comb-like structure that is present in the rest of the lamellar tissue.

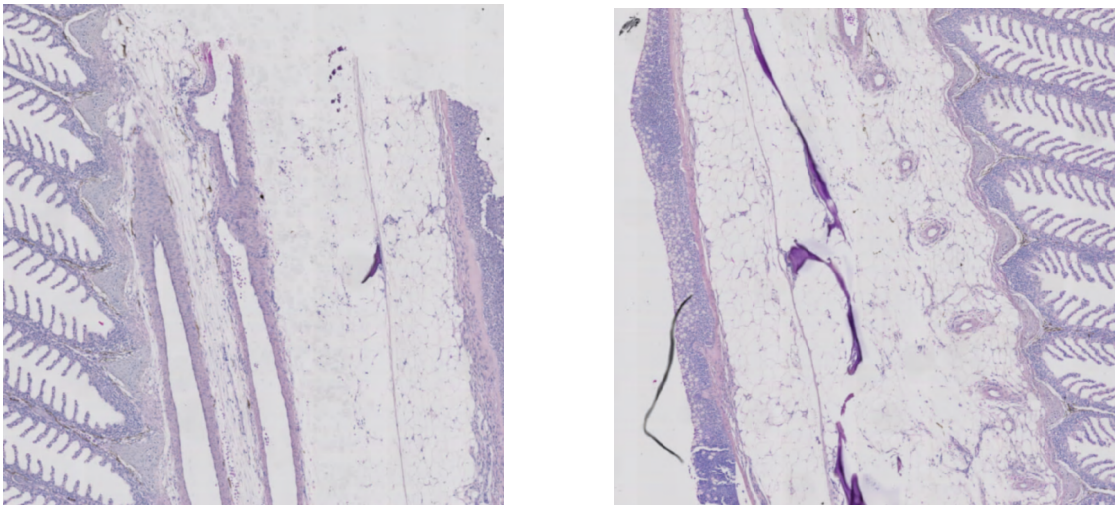
Varying quantities of excess tissue from the gill arch may also be included and erroneously classified as severe hyperplasia. Figure 7.2 shows two regions from WSIs which may be mistakenly identified as severe based on the methods developed in the previous chapter.

Given that tiles without context may be incorrectly classified, it is prudent to include another step in our pipeline to evaluate whether tiles would be included or excluded when quantifying the severity of hyperplasia in a WSI. In this chapter, we outline how anomaly detection methods, namely variational autoencoders (VAEs) can be used to classify whether regions should



(a) (b)

Figure 7.1: The tips of the primary lamellae from two different WSIs. When viewed in isolation, tiles of these regions often resemble severe hyperplasia.



(a) (b)

Figure 7.2: The base of the primary lamellae from two different WSIs. Excess tissue may often resemble severe hyperplasia.

be considered for hyperplasia analysis of lamellar tissue.

7.1 Methodology

We use techniques originally developed for the closely related task of anomaly detection, mainly because non-lamellar tissue can be deemed extraneous when it comes to the hyperplasia analysis task. Our approach involves training a neural network designed to distil an image into a compact set of parameters. Subsequently, we employ a mirrored neural network to reconstruct the original image. It's essential to note that these neural networks undergo training exclusively on lamellar tissue, which is directly relevant to hyperplasia analysis.

When an image that does not resemble lamellar tissue is passed through this model, the resulting reconstruction is noticeably less accurate. By comparing the reconstructed image with the original, we can effectively discern whether the image holds significance for the task of hyperplasia analysis. This approach allows us to eliminate tiles that may skew the results of our analysis.

7.1.1 Autoencoders

Autoencoders are a class of neural networks used for unsupervised learning tasks such as data compression, denoising, feature extraction, and in our work, anomaly detection. Combining an encoder and a decoder, autoencoders capture the underlying structure of input data by reducing its dimensionality while preserving essential features. Figure 7.4 shows this structure.

The encoder takes the original input image and maps it into a lower-dimensional representation known as the latent space. This process involves successive layers of neurons that compress and transform the input data into a compact form. The encoder essentially captures the most salient features of the input, discarding less significant information. This latent space representation serves as a compressed version of the input data. It captures the key characteristics of the data, allowing us to abstract away

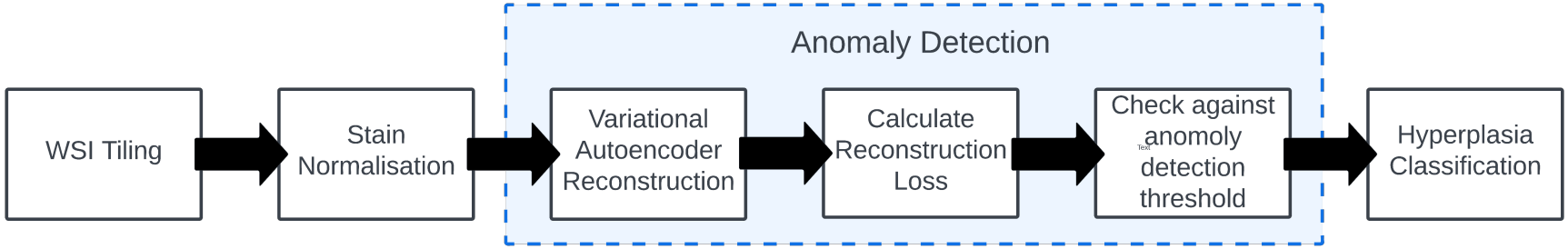


Figure 7.3: Flowchart of the anomaly detection process. After preprocessing, a variational autoencoder is used to reconstruct the image. The similarity between the original and reconstructed image is calculated using the reconstruction loss.

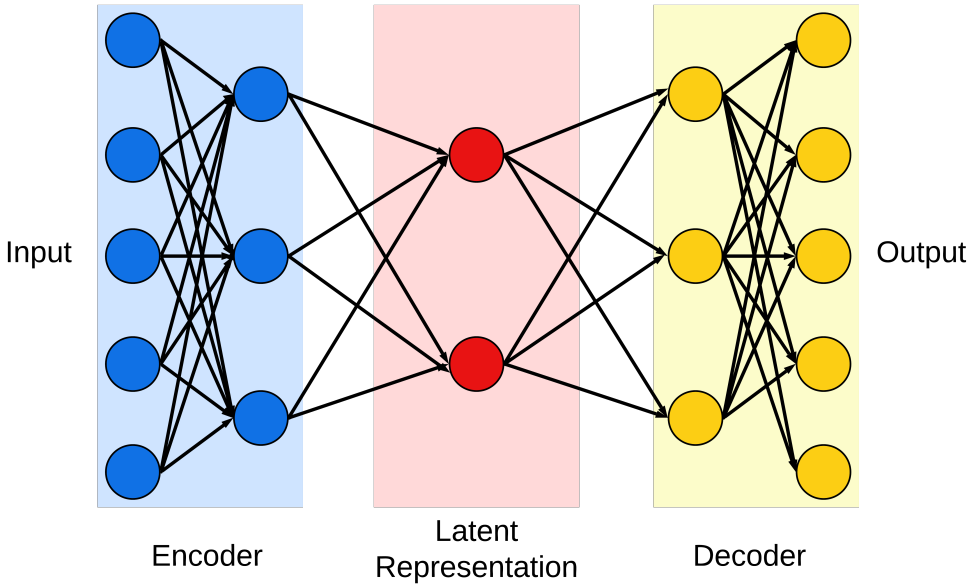


Figure 7.4: Diagram showing the encoder-decoder structure of autoencoders. The encoder takes information and reduces it to a much smaller latent representation which can then be used to recreate the infiormation using the decoder.

noise and redundancy.

The decoder then receives the latent space representation and aims to reconstruct the original image from it. Through a symmetrical architecture of the encoder, the decoder reverses the compression process, gradually expanding the latent representation into a reconstruction of the initial input. During training, the autoencoder minimises a reconstruction loss, quantifying the difference between the input image and the reconstructed output. This loss drives the network to learn representations that capture the most important characteristics of the data, leading to more effective data compression and reconstruction. Our work uses this reconstruction loss value to identify anomalous tissue, essentially appropriating the anomaly detection process.

7.1.2 Variational Autoencoders (VAEs)

Variational Autoencoders (VAEs) extend the concept of autoencoders by incorporating probabilistic elements, enabling more sophisticated and versa-

tile representations of complex data. Unlike traditional autoencoders that map input data to a fixed latent space point, VAEs map inputs to probability distributions in the latent space. This introduces uncertainty and flexibility, allowing the model to account for the inherent variability in real-world data. To facilitate gradient-based training, VAEs employ a reparameterisation trick. Instead of directly sampling from the distribution produced by the encoder, the model samples from a simpler distribution (typically Gaussian) and then transforms the samples using the encoder's learned parameters. This enables efficient gradient propagation during training. This is described in Formula 7.1, where a stochastic latent variable, z, within a standard Gaussian distribution, $\epsilon \sim \mathcal{N}(0, 1)$ is sampled for each parametric feature of the distribution, e.g. μ and σ, as predicted by the encoder.

$$\begin{aligned} z \sim q_{\mu,\sigma}(z) &= \mathcal{N}(\mu, \sigma^2) \\ \epsilon &\sim \mathcal{N}(0, 1) \\ z &= \mu + \epsilon \cdot \sigma \end{aligned} \tag{7.1}$$

VAEs introduce a regularisation term in the form of the Kullback-Leibler (KL) divergence for model training. This term encourages the learned latent space to conform to a chosen prior distribution. This is described in Formula 7.2, where the divergence of the distributions, P and Q, can be thought of as the similarity of the two distributions obtained from the original image, and the VAE reconstructed image.

$$D_{\mathrm{KL}}(P \| Q) = \sum_{x} P(x) \log \left(\frac{P(x)}{Q(x)} \right) \tag{7.2}$$

7.1.3 Reconstruction Loss

A comparison between the initial image, and the reconstructed image is performed to evaluate the accuracy of the reconstruction. In this work, we use binary cross-entropy loss to do this:

$$\textit{Binary Cross-Entropy Loss} = -(y \log(p) + (1 - y) \log(1 - p)) \tag{7.3}$$

Binary Cross-Entropy Loss is a commonly used loss function in machine

learning for binary classification problems. It measures the dissimilarity between predicted probabilities and actual binary labels (0 or 1) and is particularly well-suited for problems where the goal is to predict one of two mutually exclusive classes.

As mentioned previously, we can reuse the reconstruction loss value to predict whether an image is a lamellar tissue relevant for hyperplasia analysis. We make the crucial assumption that tissue that does not exhibit the characteristics of lamellar regions will be more difficult to reconstruct and, as such, will produce a higher reconstruction loss value.

7.1.4 Model Training

We used transfer learning to train our VAE on the same training set described in the previous chapter for hyperplasia classification. The encoder consisted of a ResNet50 [128] baseline pretrained on ImageNet database [129]. The decoder had the reverse structure. The model was trained over 50 epochs, with a batch-size of 64, and a learning rate of 0.001. Images went through Reinhard stain normalisation [75] before being resized to 224x224, as is standard for ResNet models. The encoder generates a latent representation with 256 dimensions, which serves as the input for the decoder.

7.2 Results and Visualisation

Using the dataset described in Chapter 5 for anomaly detection, we passed both lamellar and anomaly images through our model and plotted the reconstruction loss values in the histogram shown in Figure 7.5. The histogram clearly shows that the classes are separable.

7.2.1 Otsu Thresholding on Reconstruction Losses

In our analysis, we employ Otsu thresholding to automatically determine an optimal threshold value for distinguishing between two classes within the data. The Otsu method, as depicted in Formula 7.4, systematically evaluates a range of potential thresholds (t) to identify the threshold that minimises

the intra-class variance. A random subset of 200 was used to systematically calculate the threshold value. 100 tiles exhibiting anomolous/irrelevant were manually selected along with 100 tiles from the training set, with each hyperplasia severity classification equally represented.

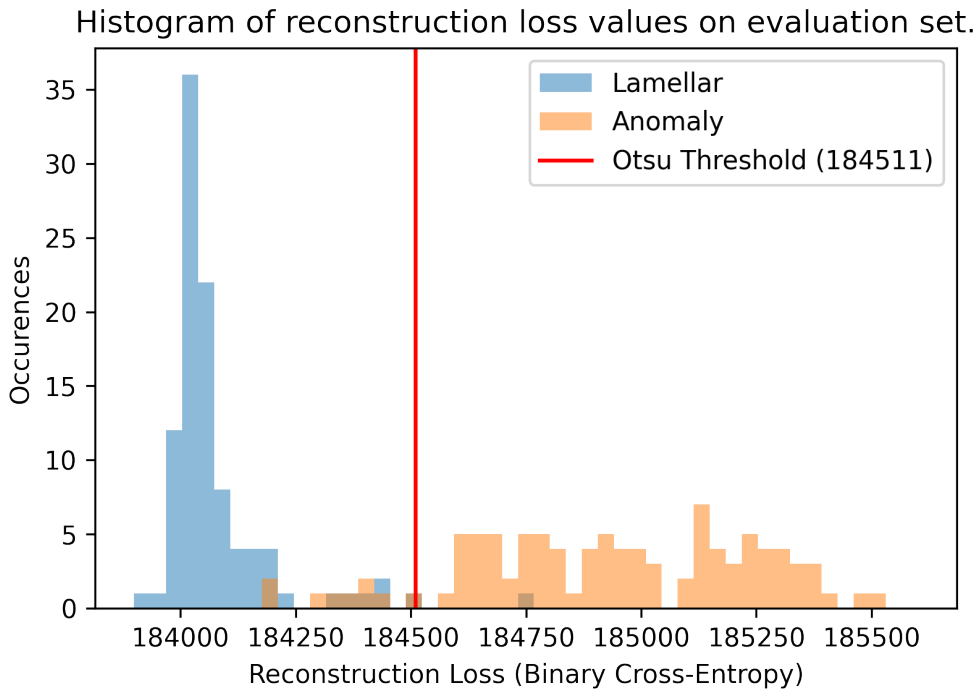


Figure 7.5: A histogram of the anomaly and lamellae class reconstruction losses, with the value obtained from Otsu thresholding shown in red.

The intra-class variance, denoted as $\sigma_w^2(t)$, is calculated as a weighted sum of the variances associated with the two classes. These weights are represented by ω_0 and ω_1, which respectively correspond to the probabilities of occurrence of the two classes. The Otsu method effectively searches for the threshold value that optimally separates these classes by minimising the variance within each class.

$$\sigma_w^2(t) = \omega_0(t)\sigma_0^2(t) + \omega_1(t)\sigma_1^2(t) \tag{7.4}$$

To demonstrate the effectiveness of this approach, we generated the visualisations shown in Figure 7.6 b, and c. Our anomaly detection approach has successfully identified some regions of tissue that are abnormal for the task of hyperplasia analysis, however, it has still failed to identify some of

the regions classified as severe at the tips and base of the gill arch.

7.3 Conclusions

Our approach has demonstrated the ability to accurately identify certain abnormal tissue regions that may negatively impact the task of hyperplasia analysis, allowing us to disregard them when considering the global slide hyperplasia score.

Despite these promising results, it has still failed to detect anomalous regions classified as severe at the tips and base of the gill arch. These areas pose a unique challenge due to their similarity to severe hyperplasia, and our approach has encountered difficulty distinguishing them accurately.

These preliminary results underscore the need for further refinement and exploration in our anomaly detection methodology. Future work may consider other neural networks for the VAE, other methods of threshold analysis, and a larger dataset for analysing which threshold value is appropriate for this task.

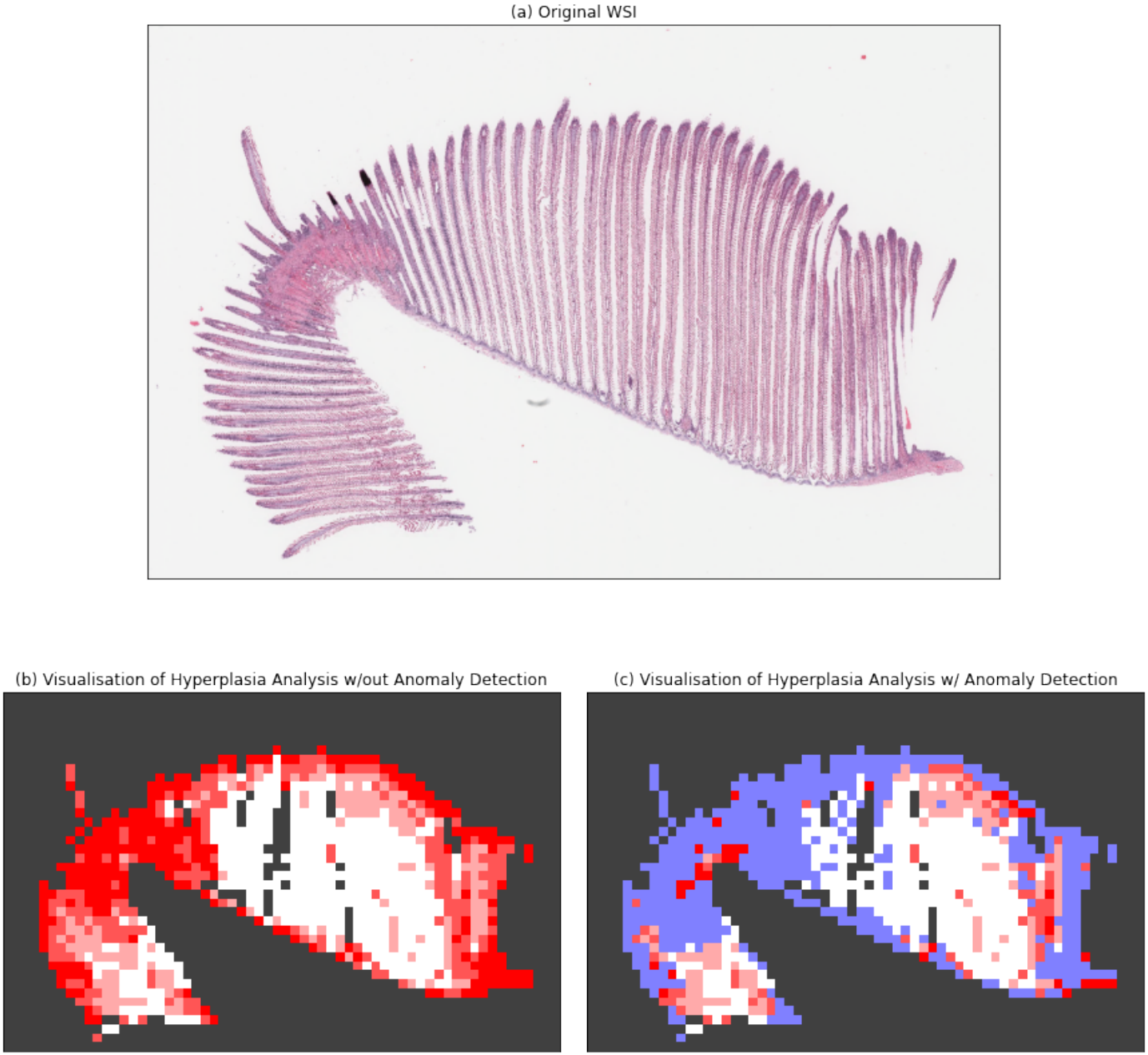


Figure 7.6: Shows a predominantly healthy gill image and two visualisations based on our method. The first shows our approach without anomaly detection, and the second represents anomalous regions in blue.

8 Lesion Score Aggregation in Gill Histology Images

Up to this point, the assessment of our model's performance has focused on the individual tile level. However, in this chapter, we focus on evaluating the model's capability to generate global hyperplasia scores for entire WSIs. We explore various techniques for aggregating the tile-level information and examine how closely these aggregated scores align with those generated by histopathologists.

8.1 Pipeline Overview

In this chapter, we evaluate the effectiveness of components of our pipeleine. The three main stages under consideration are anomaly removal, hyperplasia classification models, and metric aggregation methods. Some blocks of our pipeline remain consistent across all experiments, such as using Reinhard stain normalisation and preprocessing methods and removing tiles lacking sufficient tissue for analysis. Figure 8.1 shows a flowchart of our pipeline.

8.2 Region of Interest Classification

In our analysis, we begin by discarding empty tiles and those with insufficient tissue area for meaningful analysis. We then employ the ResNet50-based Variational Autoencoder (VAE), previously developed and trained in Chapter 7. After reconstructing a tile, we calculate the reconstruction loss

Figure 8.1: The full hyperplasia analysis pipeline for Gill WSIs.

and compare it to the same Otsu-based threshold value to differentiate between anomalous and lamellar tiles.

8.3 Hyperplasia Classification Models

Tiles which have been deemed relevant are then considered for hyperplasia classification. We employ the two classification models trained and evaluated in Chapter 6: our EWT-LP approach and ResNet18.

8.4 Metric Aggregation

Tiles in our analysis fall into one of four categories: normal, mild, moderate, or severe. Consequently, we explore three distinct methods for condensing this information into a singular metric: mean averaging of scores, modal aggregation, and median average.

In our 20 gill Whole Slide Images (WSIs) dataset, each has been assigned a hyperplasia score ranging from 0 to 3. Here, 0 signifies the absence of hyperplasia, while 3 indicates severe hyperplasia. Under the modal aggregation metric, we link a WSI's most frequently occurring severity class to its respective hyperplasia score. The number of WSIs associated with each class is shown in Figure 5.5.

In the case of mean score averaging, we start by assigning each tile its corresponding hyperplasia score value, as with the process used in modal aggregation. Subsequently, we calculate the mean average of these hyperplasia scores. We round our final score to the nearest whole integer to facilitate a direct comparison with the ground truth dataset scores.

8.5 Results

For each of the 20 Whole Slide Images (WSIs), predictions were generated using mean, modal, and median averaging, as presented in Table 8.1. The colours within Table 8.1 correspond to the level of agreement between the

developed methods and the expert score, where green signifies congruence, yellow indicates a one-degree disagreement, orange represents a two-degree disagreement, and red signifies complete divergence. Additional visualisations, both with and without the Variational Autoencoder (VAE) component, can be found in Appendices 1 and 2. These appendices also include bar charts illustrating the number of tiles associated with each class, both before and after incorporating the VAE component.

Overall, our pipeline has yielded mixed results. The absence of congruence in most cases can be attributed to several factors. Firstly, the classification component sometimes predicts tiles inaccurately. Additionally, relying on a threshold value for the reconstruction loss to eliminate non-lamellar tissue may introduce weaknesses. The use of averaging techniques can lead to a drift in the global metric towards mild and moderate scores. Lastly, differences may exist in the approaches employed by our method and expert histologists when summarising hyperplasia into a single metric. While an expert may arrive at a final score based on a subjective impression shaped by their extensive experience, the metric generated by our approach is fundamentally deterministic and explainable. This is in stark contrast to the subjective scores which may be produced between different histologist from different clinical settings. Our results also highlight the weakness of summarising a characteristic such as hyperplasia in a single metric when the reality may be more complex. We will discuss and expand upon the implications of these findings in Chapter 9.

		Without Anomaly Detection						With Anomaly Detection					
		EWT-LP			ResNet18			EWT-LP			ResNet18		
ImageID	Expert Label	Mean	Mode	Median	Mean	Mode	Median	Mean	Mode	Median	Mean	Mode	Median
1	2	2	2	2	2	2	2	2	2	2	2	2	2
2	3	2	2	2	2	2	2	2	2	2	2	2	2
3	1	3	3	3	2	2	2	2	3	3	2	2	2
4	0	2	2	2	2	2	2	1	2	1	2	2	2
5	1	2	3	2	2	2	2	2	3	2	2	2	2
6	1	2	2	2	2	2	2	2	2	2	2	2	2
7	2	2	3	3	2	3	2	2	3	2	2	3	2
8	3	2	3	2	2	2	2	2	2	2	2	2	2
9	2	2	3	2	2	2	2	2	3	2	2	2	2
10	3	2	3	2	2	2	2	2	3	3	2	2	2
11	0	1	1	1	2	2	2	2	1	2	2	2	2
12	0	2	2	2	2	2	2	2	3	3	2	3	2
13	1	2	2	2	2	2	2	2	2	2	2	2	2
14	3	2	3	2	2	2	2	1	2	2	1	2	2
15	1	2	3	2	2	2	2	2	2	2	2	2	2
16	2	2	2	2	2	2	2	1	2	2	2	2	2
17	3	2	3	3	2	3	2	2	2	2	2	2	2
18	1	2	2	2	2	2	2	2	3	2	2	3	2
19	0	2	2	2	2	2	2	2	3	3	2	2	2
20	0	2	2	2	2	2	2	2	3	3	2	2	2

Table 8.1: Table showing predicted labels based on mean, mode, and median averaging. Colour represents similarity to the expert label. Green: agreement between predicted and target. Yellow: Off-by-one between predicted and target. Orange: Off-by-two between predicted and target. Red: Off-by-three.

9 Discussion

In this chapter, we provide a comprehensive overview of the key findings derived from this project, and where feasible, draw comparisons with existing research. We delve into a thorough analysis of the results presented in the preceding chapters, aiming to explain how these findings contribute to addressing the research questions posed in this thesis. Additionally, we explore both the practical and theoretical implications of this research, with a specific focus on its potential impact on policies, practices, and the direction of future research endeavours. To maintain transparency, we discuss the limitations of our work and acknowledge the weaknesses in the methods developed. Following this, we discuss prospective avenues for future research, offering recommendations informed by the challenges encountered during the course of this project.

9.1 Information Retrieval for Marine Mammal Necropsy Analysis

In Chapter 4, we assessed the effectiveness of our information retrieval framework through a use-case analysis - bottlenose dolphin attacks (BDAs) on harbour porpoises. While the deterministic classifier was able to accurately classify cases of BDA in most cases, we found a significant number of false positives associated with deficiencies in our entity-relation engine. Furthermore, the ground truth labels used would only list the ultimate cause of death, meaning that secondary contributing pathologies or lesions were not represented.

While applying text mining methods to marine mammal post-mortem reports represents a novel approach, it's important to recognise that within the broader field of natural language processing, our method may be considered less complex compared to recent advancements in text generation. Notably, some large language models (LLMs) have demonstrated the capability to automatically comprehend intricate specialised texts spanning various domains.

An intriguing avenue for future exploration lies in the potential replacement of our ontology-driven entity-relation engine with a more advanced model. This enhanced model could still retain the capacity to summarise an animal's condition in a format suitable for storage in a non-relational database—a choice that still holds intrinsic value due to its ability to facilitate swift and efficient analysis rooted in specific anatomical observations. Despite the evolving landscape of text analysis methods, there may still be inherent advantages to preserving the simplicity and accessibility of our current database structure. It should also be noted that a more sophisticated model would allow for more rapid development of similar approaches in new domains, given the time-consuming nature of developing entity recognition engines, such as the one developed in this project.

Another limitation of this work is that all the reports we analysed came from a single organisation. It's unclear how reporting methods for marine mammal strandings might differ among organisations in different countries. Differences in language and terminology could be significant, potentially affecting the applicability of the ontologies we developed to other datasets. However, our method is adaptable and can accommodate the inclusion of new synonymous terms, which may help address this limitation when working with diverse datasets from different sources.

We were also unable to demonstrate the potential of fusing information obtained from pathology reports and corresponding image data, due to the change in circumstances in data availability. It should be noted that changes in the way that organisations store data could significantly improve data accessibility, allowing for complex pathological and epidemiological analysis of populations without requiring information retrieval techniques such as those used in this project. This can be done by storing findings in an aggregable format, rather than text.

9.2 Empirical Wavelet Transforms for Hyperplasia Score Classification

We devised and assessed a feature extraction technique employing empirical wavelet transforms in conjunction with probability density functions. Our approach used Gaussian and Gaussian Mixture models to extract significant features from subband images for the training and evaluation of our neural network model. We conducted a comparative analysis of our model against several deep learning methods, revealing that our approach consistently achieved comparable, and often superior, performance in terms of both accuracy and training time. Additionally, we conducted an evaluation of three stain normalisation methods—Macenko, Reinhard, and Modified Reinhard—across all models. Among these methods, we identified that Reinhard consistently yielded the highest accuracy in our evaluations.

Several limitations are associated with our hyperplasia classification approach, primarily stemming from how we framed the problem. We chose to employ a four-class classification system, aligning it with the scoring methodology commonly used by histopathologists when assessing gill Whole Slide Images (WSIs). This system was designed to harmonise with prevalent terminology in gill condition reporting: 'normal,' 'mild,' 'moderate,' and 'severe.' However, this approach may fall short of ideal due to the discrete nature of hyperplasia, which doesn't neatly conform to categorical boundaries. Furthermore, distinguishing between mild and moderate cases is often subjective, relying heavily on the examiner's professional experience. While we developed a rubric to account for this when creating our datasets, there still exists no industry standard for tile-level analysis given that our work is the first of its kind.

While our current methodology employs a single metric to represent hyperplasia in gill WSIs, a more nuanced understanding of how the lesion presents itself is required. For instance, the existing 0-3 scoring system does not capture the distribution of hyperplasia. It fails to distinguish between focal and general hyperplasia, both of which hold substantial clinical significance. Nevertheless, given the available data, the tile level's 0-3 scoring system remains suitable.

In the future, it would be worthwhile to explore unsupervised scoring methods at the tile level on a larger dataset. Such an approach could potentially yield more comprehensive insights and make the development of aggregation metrics at the whole-slide level a more tractable task.

In our research, we centred our parametric feature generation approach around the 2D Littlewood-Paley EWT. This choice was primarily motivated by the method's ability to produce a consistent number of boundaries, facilitating the generation of uniform feature vectors. An intriguing avenue for future exploration lies in the consideration of how EWT-based approaches such as ridglet, curvelet, and tensor could be harnessed for feature extraction.

To pave the way for more efficient development, it would be prudent to initially benchmark these methods on well-established texture-based datasets before their application to medical images. This strategic approach ensures a solid foundation and a clear understanding of their performance characteristics, allowing for a more seamless transition to the domain of medical image analysis.

Another avenue worthy of future exploration pertains to the application of deep learning-based stain normalisation methods. Although the techniques employed in our study were sufficient for our preliminary work, it's worth noting that stain normalisation in medical images is a dynamic and evolving field of research. Therefore, there is considerable potential for further investigation and experimentation in this domain in future research endeavours.

9.3 Context-Aware Hyperplasia Analysis using Variational Autoencoders

Our approach demonstrated its capability to accurately identify specific abnormal tissue regions, which could potentially interfere with hyperplasia analysis, improving some test visualisations. This allowed us to exclude these regions when calculating the overall slide hyperplasia score. However, despite these promising outcomes, our approach still faces challenges

in detecting abnormal regions classified as severe, particularly at the tips and base of the gill arch. These areas present a distinct challenge due to their resemblance to severe hyperplasia, making it difficult for our method to differentiate them accurately. Due to the use of a threshold value for the reconstruction loss, some slides were over-estimated to contain anomalous tissue, and vice-versa.

The use of anomaly detection methods to classify regions of interest was to remedy the indiscriminate way in which our pipeline processes any tile containing a sufficient quantity of tissue pixels. In our project, we relied on the loss generated through the comparison of a tile with its ResNet50-based VAE reconstruction, coupled with Otsu thresholding based on a dataset equally divided between our lamellar and anomaly classes. While this approach sufficed for our preliminary analysis, it's important to note that the effectiveness of our Otsu threshold is significantly influenced by the proportion of lamellar to anomaly images in the dataset. Consequently, it does not purely represent a threshold value that delineates the divide between these two classes. This is illustrated by the visualisation in Appendices 1 and 2, which highlight how a single threshold value, effective in eliminating extraneous tissue at the base and tips in one image, may inadvertently remove essential tissue in another WSI.

Conducting a comprehensive evaluation of various anomaly detection methods, encompassing a range of autoencoders and variational autoencoders, holds the potential to foster greater alignment between the scores generated by our pipeline and those provided by histopathologists. This evaluation should involve a diverse set of neural networks, with a focus on assessing their performance while considering the trade-off between accuracy and computational runtime—a particularly pertinent consideration for computationally intensive methods.

Additionally, there exists the prospect of developing segmentation and classification models to differentiate between lamellar and non-lamellar tissue. However, this avenue was not pursued in our study due to the substantial workload associated with annotating such data. Nevertheless, it's worth noting that the detection of regions of interest using similar methods could find broader applications across the realm of general medical imaging.

9.4 Lesion Score Aggregation in Gill Histology Images

We compared various metric generation approaches for representing hyperplasia within Whole Slide Images (WSIs) and contrasted these metrics with those generated by expert histopathologists. Notably, we observed significant disparities in the scores, which could be attributed to limitations inherent in the aggregation techniques employed and the methods utilised by histopathologists to summarise severity. Interestingly, our incorporation of the VAE component exhibited mixed effects on the alignment of opinions between the pipeline and the experts. In certain cases, it facilitated a convergence of opinions; in others, it led to divergence. While this marks a promising starting point for understanding how hyperplasia can be quantified, further investigation with an expanded dataset is imperative to gain deeper insights into the underlying reasons for the discrepancies observed between the model and expert assessments.

9.5 Research Implications

At the time of writing, our work is the only project to apply data-driven approaches to marine mammal post-mortem reports and gill WSIs. Further development of these, and similar methods, would have several significant implications for industry, research, and policy.

- **Improved Understanding of Health Patterns**: Data-driven analysis plays a pivotal role in enhancing our understanding of health patterns and disease dynamics within aquatic animal populations. This approach becomes particularly valuable when dealing with large, yet previously underutilised datasets, often characterised by their unstructured nature. By applying data-driven information retrieval methodologies, we can transform these disparate reports and images into a more comprehensive model of an animal's health state at the time of death. This holistic view empowers researchers to identify critical trends, risk factors, and emerging health issues with a granularity and precision previously unattainable. Ultimately, this enriched dataset amplifies the

efficacy of epidemiological analyses, contributing significantly to our collective understanding of aquatic animal health.

- **Improved Disease Detection**: Data-driven analysis methods excel in their capacity to enable the early detection of diseases and health anomalies among aquatic animals. What sets these approaches apart is their ability to model an animal's condition at the time of death at a granular lesion level, improving upon the traditional focus solely on the pathological level. By doing so, we can capture the emergence of pathologies with heightened precision and sensitivity. This heightened sensitivity, in turn, empowers researchers and stakeholders to identify potential health issues, facilitating the implementation of more effective disease management and prevention strategies. Timely interventions enhance the welfare of aquatic populations and contribute significantly to the overall health and sustainability of aquatic ecosystems.
- **Data Integration**: The power of data-driven analysis lies in its ability to seamlessly integrate information from diverse sources, culminating in a comprehensive perspective of aquatic animal health. This integration extends beyond just traditional datasets and encompasses crucial environmental and behavioural factors, such as water quality scores and weather conditions Combining these external variables with insights extracted from unstructured data modalities creates a holistic portrait of an animal's health. This enables researchers and stakeholders to analyse the interplay between an animal's well-being and its surroundings. Such integration is invaluable in unveiling the complex relationships between aquatic animals and their environments, allowing more informed decision-making and conservation efforts.
- **Environmental Impact Assessment**: By examining the connection between aquatic animal health and environmental factors, the methodologies developed in this project can make substantial contributions to the assessment of the environmental repercussions stemming from human activities on aquatic populations. An illustration of this potential could be the application of the information retrieval framework to reports mentioning entanglements, which often culminate in marine mammal strandings. Through this approach, we can gain insights into the broader ecological implications of individual cases, shedding light on the direct and indirect consequences of human activities on aquatic

ecosystems.

- **Policy and Regulatory Decisions**: Robust data analysis can inform policy and regulatory decisions related to aquatic animal health. Evidence-based policies can enhance the sustainability and welfare of aquatic ecosystems.

10 Conclusions and Future Work

This thesis addresses the significant challenge of analysing unstructured aquatic animal health data. The overarching objective was to develop a comprehensive set of techniques encompassing text mining, signal processing, image analysis, and machine learning to handle such complex data effectively, ultimately enabling information aggregation across historic datasets.

In pursuit of these goals, we achieved notable milestones in two key areas:

Text Analysis: We devised an ontology-based framework tailored for extracting and organising information from aquatic animal post-mortem reports, with a primary focus on gross pathology reports. Initially applied to marine mammal stranding reports, this adaptable framework holds potential for various species and multiple report types.

Gill WSI Analysis: Our research led to the creation of innovative methods for identifying and analysing lesions within WSIs of Atlantic salmon gills. This approach introduced a novel feature extraction technique leveraging the empirical wavelet transform and incorporated context awareness through the use of a variational autoencoder to identify regions of interest within histology images.

Our research has identified a substantial research gap in the domain of sophisticated analysis of unstructured data modalities. Our investigations have encompassed the exploration of promising approaches for extracting information from textual post-mortem reports. Additionally, we have pioneered innovative signal and image processing techniques for the compre-

hensive analysis of lesions within gill Whole Slide Images (WSIs). We have also established that our explainable method often diverges from expert scores.

These accomplishments collectively represent a significant advancement in the analysis of unstructured aquatic animal health data. They enhance the efficiency and comprehensiveness of data processing, contributing to a deeper understanding of aquatic animal health. At the time of writing, this project stands as a unique endeavour, applying data-driven methodologies to marine mammal post-mortem reports and gill WSIs, paving the way for more sophisticated and holistic data analysis in this domain.

10.1 Future Work

This project lays the groundwork for the analysis of unstructured data modalities in aquatic animal health, opening doors to various potential areas of future research:

- **Integration of LLM-Based Approaches**: Explore the incorporation of new Large Language Models (LLMs) in tandem with ontology-driven storage methods to enhance the understanding of reports.
- **Extension to Other Species and Report Types**: Apply ontology-driven information retrieval techniques to different species and diverse report types to broaden the scope of analysis.
- **Historical Archives Analysis**: Investigate the application of information retrieval approaches to historical archives, enabling epidemiological analysis at the lesion/observation level.
- **Parametric Feature Generation with EWT**: Further develop parametric feature generation methods in conjunction with EWT. This includes exploring tensor, curvelet, and ridglet approaches, with a primary focus on benchmarking against texture analysis datasets.
- **Expansion of Hyperplasia Analysis Datasets**: Create larger hyperplasia analysis datasets encompassing a broader range of pathologies and variations in staining.
- **Exploration of Other Gill Lesions**: Investigate additional lesions within gills, such as mucous cell analysis, hypertrophy, lamellar fusion, and inflammation.

- **Extension to Other Organs**: Develop data-driven methods and datasets for other organs within fish, enabling comprehensive profiling of lesions across the entire organism.
- **Anomaly Detection Methods**: Conduct further experiments with anomaly detection methods, including a comparative analysis of normal autoencoders and Variational Autoencoders (VAEs). Explore the potential of utilising the VAE's latent space for producing a discrete representation of hyperplasia severity in tiles.
- **Extension to Other Organs**: Develop data-driven methods and datasets for other organs within fish, enabling comprehensive profiling of lesions across the entire organism.
- **EWT-based Anomaly Detection**: Using the EWT-based feature extraction methods to create both shallow and deep anomaly detection methods.

Bibliography

[1] C. A. Barbano, D. Perlo, E. Tartaglione, A. Fiandrotti, L. Bertero, P. Cassoni, and M. Grangetto, "Unitopatho, a labeled histopathological dataset for colorectal polyps classification and adenoma dysplasia grading," in *2021 IEEE International Conference on Image Processing (ICIP)*. IEEE, 2021, pp. 76–80.

[2] C. A. Barbano and A. Pedersen, "Eidoslab/torchstain: v1.2.0-stable," Aug. 2022. [Online]. Available: https://doi.org/10.5281/zenodo.6979540

[3] A. Carmichael, D. Bhowmik, J. Baily, A. Brownlow, G. J. Gunn, and A. Reeves, "Ir-man: An information retrieval framework for marine animal necropsy analysis," in *11th ACM International Conference on Bioinformatics, Computational Biology and Health Informatics (BCB'20)*. Association for Computing Machinery (ACM), 2020.

[4] A. F. Carmichael, J. L. Baily, A. Reeves, G. Ochoa, A. S. Boerlage, G. Gunn, R. Allshire, and D. Bhowmik, "Analysing hyperplasia in Atlantic salmon gills using empirical wavelets," in *Medical Imaging 2023: Digital and Computational Pathology*, vol. 12471. SPIE, 2023, pp. 116–123.

[5] M. Shehab, L. Abualigah, Q. Shambour, M. A. Abu-Hashem, M. K. Y. Shambour, A. I. Alsalibi, and A. H. Gandomi, "Machine learning in medical applications: A review of state-of-the-art methods," *Computers in Biology and Medicine*, vol. 145, p. 105458, 2022.

[6] D. Anderson, M. V. Bjarnadottir, and Z. Nenova, "Machine learning in healthcare: Operational and financial impact," *Innovative Technology at the Interface of Finance and Operations: Volume I*, pp. 153–174, 2022.

[7] P. S. Basran and R. B. Appleby, "The unmet potential of artificial intelligence in veterinary medicine," *American Journal of Veterinary Research*, vol. 83, no. 5, pp. 385–392, 2022.

[8] R. S. Williams, D. J. Curnick, J. L. Barber, A. Brownlow, N. J. Davison, R. Deaville, M. Perkins, S. Jobling, and P. D. Jepson, "Juvenile harbor porpoises in the UK are exposed to a more neurotoxic mixture of polychlorinated biphenyls than adults," *Science of the Total Environment*, vol. 708, p. 134835, 2020.

[9] R. Williams, M. t. Doeschate, D. J. Curnick, A. Brownlow, J. L. Barber, N. J. Davison, R. Deaville, M. Perkins, P. D. Jepson, and S. Jobling, "Levels of polychlorinated biphenyls are still associated with toxic effects in harbor porpoises (phocoena phocoena) despite having fallen below proposed toxicity thresholds," *Environmental Science & Technology*, vol. 54, no. 4, pp. 2277–2286, 2020.

[10] S. E. Nelms, J. Barnett, A. Brownlow, N. Davison, R. Deaville, T. S. Galloway, P. K. Lindeque, D. Santillo, and B. J. Godley, "Microplastics in marine mammals stranded around the British coast: ubiquitous but transitory?" *Scientific Reports*, vol. 9, no. 1, pp. 1–8, 2019.

[11] I. Lozano-Muñoz, J. Wacyk, C. Kretschmer, Y. Vásquez-Martínez, and M. Cortez-San Martin, "Antimicrobial resistance in Chilean marine-farmed salmon: Improving food safety through one health," *One Health*, vol. 12, p. 100219, 2021.

[12] J.-E. Dessen, T.-k. K. Østbye, B. Ruyter, M. Bou, M. Thomassen, and K.-A. Rørvik, "Sudden increased mortality in large seemingly healthy farmed Atlantic salmon (salmo salar l.) was associated with environmental and dietary changes," *Journal of Applied Aquaculture*, vol. 33, no. 2, pp. 165–182, 2021.

[13] R. Leaper, E. MacLennan, A. Brownlow, S. V. Calderan, K. Dyke, P. G. Evans, L. Hartny-Mills, D. Jarvis, L. McWhinnie, A. Philp *et al.*, "Estimates of humpback and minke whale entanglements in the Scottish static pot (creel) fishery," *Endangered Species Research*, vol. 49, pp. 217–232, 2022.

[14] A. Ahmad, S. B. Kurniawan, S. R. S. Abdullah, A. R. Othman, and H. A. Hasan, "Contaminants of emerging concern (cecs) in aquaculture effluent: Insight into breeding and rearing activities, alarming impacts, regulations, performance of wastewater treatment unit and future approaches," *Chemosphere*, vol. 290, p. 133319, 2022.

[15] J. A. Pradeepkiran, "Aquaculture role in global food security with nutritional value: a review," *Translational Animal Science*, vol. 3, no. 2, pp. 903–910, 2019.

[16] Food and A. O. of the United Nations. (2022) The state of world fisheries and aquaculture 2022. 25/09/2023. [Online]. Available: https://www.fao.org/3/cc0461en/online/sofia/2022/world-fisheries-aquaculture.html#box1

[17] J. B. West, "Galen and the beginnings of western physiology," *American Journal of Physiology-Lung Cellular and Molecular Physiology*, vol. 307, no. 2, pp. L121–L128, 2014.

[18] D. Lanjewar, "Autopsy: Introduction," *Autopsy Practices*, p. 1, 2021.

[19] A. J. Paul, "The need and status of sea turtle conservation and survey of associated computer vision advances," in *2021 IEEE 8th Uttar Pradesh Section International Conference on Electrical, Electronics and Computer Engineering (UPCON)*, 2021, pp. 1–8.

[20] E. N. Rodofili, V. Lecours, and M. LaRue, "Remote sensing techniques for automated marine mammals detection: a review of methods and current challenges," *PeerJ*, vol. 10, p. e13540, 2022.

[21] A. Coram, N. A. S. Abreo, R. P. Ellis, and K. F. Thompson, "Contribution of social media to cetacean research in Southeast Asia: illuminating populations vulnerable to litter," *Biodiversity and Conservation*, vol. 30, no. 8-9, pp. 2341–2359, 2021.

[22] C. Peter, P. L. K. Mustika, J. M. V. Acebes, N. Chansue, L. Dolar, G. S. Ham, E. Hines, W. Hte, G. Minton, L. S. Ponnampalam *et al.*, "Commentary on coram et al.(2021) on the use of facebook to understand marine mammal stranding issues in Southeast Asia," *Biodiversity and Conservation*, vol. 31, no. 7, pp. 1987–1994, 2022.

[23] "Scotlandapos;s Aquaculture | Our Aquaculture — aquaculture.scotland.gov.uk," http://aquaculture.scotland.gov.uk/our_aquaculture/our_aquaculture.aspx.

[24] "Aquaculture - gov.scot — gov.scot," https://www.gov.scot/policies/aquaculture/, [Accessed 03-May-2023].

[25] Q. Jiang, N. Bhattarai, M. Pahlow, and Z. Xu, “Environmental sustainability and footprints of global aquaculture,” *Resources, Conservation and Recycling*, vol. 180, p. 106183, 2022.

[26] A. S. Boerlage, A. Ashby, A. Herrero, A. Reeves, G. J. Gunn, and H. D. Rodger, “Epidemiology of marine gill diseases in Atlantic salmon (Salmo salar) aquaculture: a review,” *Reviews in aquaculture*, vol. 12, no. 4, pp. 2140–2159, 2020.

[27] S. Maulu, O. J. Hasimuna, L. H. Haambiya, C. Monde, C. G. Musuka, T. H. Makorwa, B. P. Munganga, K. J. Phiri, and J. D. Nsekanabo, “Climate change effects on aquaculture production: sustainability implications, mitigation, and adaptations,” *Frontiers in Sustainable Food Systems*, vol. 5, p. 609097, 2021.

[28] S. B. Schumacker, M. Davis, T. Fote, S. Meredith Moore, K. R. Moreland, J. Runnebaum, P. Sullivan, and R. Yamada, “Workforce development and critical components for developing a modern fisheries, aquaculture, and seafood workforce,” 2022.

[29] D. Komura and S. Ishikawa, “Machine learning methods for histopathological image analysis,” *Computational and structural biotechnology journal*, vol. 16, pp. 34–42, 2018.

[30] A. Rohani, M. Taki, and G. Bahrami, “Application of artificial intelligence for separation of live and dead rainbow trout fish eggs,” *Artificial Intelligence in Agriculture*, vol. 1, pp. 27–34, 2019.

[31] M. Saberioon, A. Gholizadeh, P. Cisar, A. Pautsina, and J. Urban, “Application of machine vision systems in aquaculture with emphasis on fish: state-of-the-art and key issues,” *Reviews in Aquaculture*, vol. 9, no. 4, pp. 369–387, 2017.

[32] A. F. Fernandes, E. M. Turra, E. R. de Alvarenga, T. L. Passafaro, F. B. Lopes, G. F. Alves, V. Singh, and G. J. Rosa, “Deep learning image segmentation for extraction of fish body measurements and prediction of body weight and carcass traits in nile tilapia,” *Computers and electronics in agriculture*, vol. 170, p. 105274, 2020.

[33] H. Zhang and F. Gui, “The application and research of new digital technology in marine aquaculture,” *Journal of Marine Science and Engineering*, vol. 11, no. 2, p. 401, 2023.

[34] G. Kaur, N. Adhikari, S. Krishnapriya, S. G. Wawale, R. Malik, A. S. Zamani, J. Perez-Falcon, J. Osei-Owusu *et al.*, “Recent advancements in deep learning frameworks for precision fish farming opportunities, challenges, and applications,” *Journal of Food Quality*, vol. 2023, 2023.

[35] H. Liu, X. Ma, Y. Yu, L. Wang, and L. Hao, “Application of deep learning-based object detection techniques in fish aquaculture: A review,” *Journal of Marine Science and Engineering*, vol. 11, no. 4, p. 867, 2023.

[36] D. Li and L. Du, “Recent advances of deep learning algorithms for aquacultural machine vision systems with emphasis on fish,” *Artificial Intelligence Review*, pp. 1–40, 2022.

[37] K. Malde, N. O. Handegard, L. Eikvil, and A.-B. Salberg, “Machine intelligence and the data-driven future of marine science,” *ICES Journal of Marine Science*, vol. 77, no. 4, pp. 1274–1285, 2020.

[38] C. Beyan and H. I. Browman, “Setting the stage for the machine intelligence era in marine science,” *ICES Journal of Marine Science*, vol. 77, no. 4, pp. 1267–1273, 2020.

[39] S. Bird, E. Klein, and E. Loper, *Natural language processing with Python: analyzing text with the natural language toolkit*. " O’Reilly Media, Inc.", 2009.

[40] J. Gilles, G. Tran, and S. Osher, “2d empirical transforms. wavelets, ridgelets, and curvelets revisited,” *SIAM Journal on Imaging Sciences*, vol. 7, no. 1, pp. 157–186, 2014.

[41] W. W. Chapman, W. Bridewell, P. Hanbury, G. F. Cooper, and B. G. Buchanan, “A simple algorithm for identifying negated findings and diseases in discharge summaries,” *Journal of biomedical informatics*, vol. 34, no. 5, pp. 301–310, 2001.

[42] H. Gao, E. J. A. Bowles, D. Carrell, and D. S. Buist, “Using natural language processing to extract mammographic findings,” *Journal of biomedical informatics*, vol. 54, pp. 77–84, 2015.

[43] J. Friedlin, M. Mahoui, J. Jones, and P. Jamieson, “Knowledge discovery and data mining of free text radiology reports,” in *IEEE First International Conference on Healthcare Informatics, Imaging and Systems Biology*. IEEE, 2011, pp. 89–96.

[44] A. Comelli, L. Agnello, and S. Vitabile, “An ontology-based retrieval system for mammographic reports,” in *IEEE Symposium on Computers and Communication (ISCC)*. IEEE, 2015, pp. 1001–1006.

[45] L. Gong, R. Yan, Q. Liu, H. Yang, G. Yang, and K. Jiang, “Extraction of biomedical informtion related to breast cancer using text mining,” in *12th International Conference on Natural Computation, Fuzzy Systems and Knowledge Discovery (ICNC-FSKD)*. IEEE, 2016, pp. 801–805.

[46] P. Sudeshna, S. Bhanumathi, and M. A. Hamlin, “Identifying symptoms and treatment for heart disease from biomedical literature using text data mining,” in *International Conference on Computation of Power, Energy Information and Commuincation (ICCPEIC)*. IEEE, 2017, pp. 170–174.

[47] S. Zhao, M. Jiang, M. Liu, B. Qin, and T. Liu, “Causaltriad: toward pseudo causal relation discovery and hypotheses generation from medical text data,” in *Proceedings of the 2018 ACM International Conference on Bioinformatics, Computational Biology, and Health Informatics*, 2018, pp. 184–193.

[48] Y. Yang, Y. Cai, W. Luo, Z. Li, Z. Ma, X. Yu, and H. Yu, “An ontology-based approach for text mining of stroke electronic medical records,” in *IEEE International Conference on Bioinformatics and Biomedicine*. IEEE, 2013, pp. 288–291.

[49] Z. Gero and J. C. Ho, “Namedkeys: Unsupervised keyphrase extraction for biomedical documents,” in *Proceedings of the 10th ACM International Conference on Bioinformatics, Computational Biology and Health Informatics*, 2019, pp. 328–337.

[50] S. B. Navathe, “Text mining and ontology applications in bioinformatics and gis,” in *Sixth International Conference on Machine Learning and Applications (ICMLA 2007)*. IEEE, 2007, pp. xviii–xix.

[51] O. Bodenreider, "The unified medical language system (umls): integrating biomedical terminology," *Nucleic acids research*, vol. 32 Database issue, pp. D267–70, 2004.

[52] V. Mala and D. Lobiyal, "Concepts extraction for medical documents using ontology," in *International Conference on Advances in Computer Engineering and Applications*. IEEE, 2015, pp. 773–777.

[53] L. Gong, J. Yan, J. Feng, and R. Yang, "A dictionary-based approach to identify biomedical concepts," in *12th International Conference on Fuzzy Systems and Knowledge Discovery (FSKD)*. IEEE, 2015, pp. 1091–1095.

[54] S. Mate, F. Köpcke, D. Toddenroth, M. Martin, H.-U. Prokosch, T. Bürkle, and T. Ganslandt, "Ontology-based data integration between clinical and research systems," *PloS one*, vol. 10, no. 1, 2015.

[55] D. Kim, J. Lee, C. H. So, H. Jeon, M. Jeong, Y. Choi, W. Yoon, M. Sung, and J. Kang, "A neural named entity recognition and multi-type normalization tool for biomedical text mining," *IEEE Access*, vol. 7, pp. 73 729–73 740, 2019.

[56] H. Wei, M. Gao, A. Zhou, F. Chen, W. Qu, C. Wang, and M. Lu, "Named entity recognition from biomedical texts using a fusion attention-based bilstm-crf," *IEEE Access*, vol. 7, pp. 73 627–73 636, 2019.

[57] N. Bollig, L. Clarke, E. Elsmo, and M. Craven, "Machine learning for syndromic surveillance using veterinary necropsy reports," *PloS one*, vol. 15, no. 2, p. e0228105, 2020.

[58] L. Furrer, S. Küker, J. Berezowski, H. Posthaus, F. Vial, F. Rinaldi, T. Poibeau, and P. Faber, "Constructing a syndromic terminology resource for veterinary text mining," 2015.

[59] S. Küker, C. Faverjon, L. Furrer, J. Berezowski, H. Posthaus, F. Rinaldi, and F. Vial, "The value of necropsy reports for animal health surveillance," *BMC veterinary research*, vol. 14, no. 1, p. 191, 2018.

[60] OpenAI. (2022) Introducing ChatGPT. [Online]. Available: https://openai.com/blog/chatgpt

[61] ——. (2022) New GPT-3 capabilities: Edit insert. [Online]. Available: https://openai.com/blog/gpt-3-edit-insert

[62] Sundar Pichai. (2023) An important next step on our AI journey. [Online]. Available: https://blog.google/technology/ai/bard-google-ai-search-updates/

[63] Meta AI. (2023) Introducing LLaMA: A foundational, 65-billion-parameter large language model. [Online]. Available: https://ai.meta.com/blog/large-language-model-llama-meta-ai/

[64] V. Liévin, C. E. Hother, and O. Winther, "Can large language models reason about medical questions?" *arXiv preprint arXiv:2207.08143*, 2022.

[65] M. Soleymani, D. Garcia, B. Jou, B. Schuller, S.-F. Chang, and M. Pantic, "A survey of multimodal sentiment analysis," *Image and Vision Computing*, vol. 65, pp. 3–14, 2017.

[66] N. Audebert, C. Herold, K. Slimani, and C. Vidal, "Multimodal deep networks for text and image-based document classification," in *Joint European Conference on Machine Learning and Knowledge Discovery in Databases*. Springer, 2019, pp. 427–443.

[67] M. Kerroumi, O. Sayem, and A. Shabou, "Visualwordgrid: Information extraction from scanned documents using a multimodal approach," *arXiv preprint arXiv:2010.02358*, 2020.

[68] J. Ferrando, J. L. Domínguez, J. Torres, R. García, D. García, D. Garrido, J. Cortada, and M. Valero, "Improving accuracy and speeding up document image classification through parallel systems," in *International Conference on Computational Science*. Springer, 2020, pp. 387–400.

[69] A. Kay, "Tesseract: an open-source optical character recognition engine," *Linux Journal*, vol. 2007, no. 159, p. 2, 2007.

[70] O. Pelka, S. Koitka, J. Rückert, F. Nensa, and C. M. Friedrich, "Radiology objects in context (roco): A multimodal image dataset," in *Intravascular Imaging and Computer Assisted Stenting and Large-Scale Annotation of Biomedical Data and Expert Label Synthesis*. Springer, 2018, pp. 180–189.

[71] X. Wang, Z. Guo, Y. Zhang, and J. Li, “Medical image labelling and semantic understanding for clinical applications,” in *International Conference of the Cross-Language Evaluation Forum for European Languages*. Springer, 2019, pp. 260–270.

[72] S. Maksoud, A. Wiliem, and B. C. Lovell, “Recurrent attention networks for medical concept prediction.” in *CLEF (Working Notes)*, 2019.

[73] R. Sonker, A. Mishra, P. Bansal, and A. Pattnaik, “Techniques for medical concept detection from multi-modal images,” in *CLEF2020 Working Notes. CEUR Workshop Proceedings, CEUR-WS. org, Thessaloniki, Greece (September 22-25 2020)*.

[74] A. Vahadane, T. Peng, S. Albarqouni, M. Baust, K. Steiger, A. M. Schlitter, A. Sethi, I. Esposito, and N. Navab, “Structure-preserved color normalization for histological images,” in *2015 IEEE 12th International Symposium on Biomedical Imaging (ISBI)*, 2015, pp. 1012–1015.

[75] E. Reinhard, M. Adhikhmin, B. Gooch, and P. Shirley, “Color transfer between images,” *IEEE Computer Graphics and Applications*, vol. 21, no. 5, pp. 34–41, 2001.

[76] M. Macenko, M. Niethammer, J. S. Marron, D. Borland, J. T. Woosley, Xiaojun Guan, C. Schmitt, and N. E. Thomas, “A method for normalizing histology slides for quantitative analysis,” in *2009 IEEE International Symposium on Biomedical Imaging: From Nano to Macro*, 2009, pp. 1107–1110.

[77] T. Araújo, G. Aresta, E. Castro, J. Rouco, P. Aguiar, C. Eloy, A. Polónia, and A. Campilho, “Classification of breast cancer histology images using convolutional neural networks,” *PloS one*, vol. 12, no. 6, p. e0177544, 2017.

[78] F. Wassmer, V. De Luca, R. Casanova, A. Soltermann, and G. Székely, “Automatic segmentation of lung carcinoma in histological images using a visual dictionary,” in *2016 IEEE 13th International Symposium on Biomedical Imaging (ISBI)*, 2016, pp. 1037–1040.

[79] S. Graham, S. A. Khurram, M. Shaban, T. Qaiser, and N. Rajpoot, “Classification of lung cancer histology images using patch-level summary statistics,” 03 2018, p. 44.

[80] S. Wang, A. Chen, L. Yang, L. Cai, Y. Xie, J. Fujimoto, A. Gazdar, and G. Xiao, "Comprehensive analysis of lung cancer pathology images to discover tumor shape and boundary features that predict survival outcome," *Scientific reports*, vol. 8, no. 1, pp. 1–9, 2018.

[81] K.-H. Yu, C. Zhang, G. J. Berry, R. B. Altman, C. Ré, D. L. Rubin, and M. Snyder, "Predicting non-small cell lung cancer prognosis by fully automated microscopic pathology image features," *Nature communications*, vol. 7, no. 1, pp. 1–10, 2016.

[82] O. S. Al-Kadi, "A gabor filter texture analysis approach for histopathological brain tumor subtype discrimination," *arXiv preprint arXiv:1704.05122*, 2017.

[83] D. Kolhatkar and N. Wankhade, "Detection and counting of blood cells using image segmentation: A review," in *2016 World Conference on Futuristic Trends in Research and Innovation for Social Welfare (Startup Conclave)*, 2016, pp. 1–5.

[84] N. Dimitriou, O. Arandjelović, and P. D. Caie, "Deep learning for whole slide image analysis: An overview," *Frontiers in Medicine*, vol. 6, 2019.

[85] D. Shen, G. Wu, and H.-I. Suk, "Deep learning in medical image analysis," *Annual review of biomedical engineering*, vol. 19, pp. 221–248, 2017.

[86] G. Litjens, T. Kooi, B. E. Bejnordi, A. A. A. Setio, F. Ciompi, M. Ghafoorian, J. A. Van Der Laak, B. Van Ginneken, and C. I. Sánchez, "A survey on deep learning in medical image analysis," *Medical image analysis*, vol. 42, pp. 60–88, 2017.

[87] J. N. Kather, C.-A. Weis, F. Bianconi, S. M. Melchers, L. R. Schad, T. Gaiser, A. Marx, and F. G. Zöllner, "Multi-class texture analysis in colorectal cancer histology," *Scientific reports*, vol. 6, p. 27988, 2016.

[88] S. Zhao, S. Zhang, J. Liu, H. Wang, J. Zhu, D. Li, and R. Zhao, "Application of machine learning in intelligent fish aquaculture: A review," *Aquaculture*, p. 736724, 2021.

[89] "Novel method for quantifying salmonid mucous cells," *Journal of fish diseases*, vol. 34, no. 12, pp. 931–936, 2011.

[90] S. Haddeland, C. C. Lazado, G. V. Merkin, O. J. Myre, M. A. Okubamichael, L.-F. Pedersen, and K. Pittman, "Dynamic morphometrics of mucous cells reveal the minimal impact of therapeutic doses of peracetic acid on Atlantic salmon gill health," *Aquaculture*, vol. 534, p. 736315, 2021.

[91] N. S. Jayasuriya, "Gill image analysis : a tool for assessing pathophysiological and morphometric changes in the gill of atlantic salmon (salmo salar l.)," Ph.D. dissertation, University of Stirling, 2014.

[92] L. Sveen, G. Timmerhaus, L.-H. Johansen, and E. Ytteborg, "Deep neural network analysis-a paradigm shift for histological examination of health and welfare of farmed fish," *Aquaculture*, vol. 532, p. 736024, 2021.

[93] A. H. Sweidan, N. El-Bendary, A. E. Hassanien, O. M. Hegazy, and A. E. Mohamed, "Machine learning based approach for water pollution detection via fish liver microscopic images analysis," in *2014 9th International Conference on Computer Engineering Systems (ICCES)*, 2014, pp. 253–258.

[94] ——, "Water quality classification approach based on bio-inspired gray wolf optimization," in *2015 7th International Conference of Soft Computing and Pattern Recognition (SoCPaR)*, 2015, pp. 1–6.

[95] P. F. Silva, C. McGurk, K. Thompson, N. Jayasuriya, and J. Bron, "Development of a quantitative semi-automated system for intestinal morphology assessment in Atlantic salmon, using image analysis," *Aquaculture*, vol. 442, pp. 100–111, 2015.

[96] N. E. Huang, Z. Shen, S. R. Long, M. C. Wu, H. H. Shih, Q. Zheng, N.-C. Yen, C. C. Tung, and H. H. Liu, "The empirical mode decomposition and the hilbert spectrum for nonlinear and non-stationary time series analysis," *Proceedings of the Royal Society of London. Series A: mathematical, physical and engineering sciences*, vol. 454, no. 1971, pp. 903–995, 1998.

[97] J. Gilles, "Empirical wavelet transform," *IEEE Transactions on Signal Processing*, vol. 61, no. 16, pp. 3999–4010, 2013.

[98] S. Maheshwari, R. B. Pachori, and U. R. Acharya, "Automated diagnosis of glaucoma using empirical wavelet transform and correntropy features extracted from fundus images," *IEEE journal of biomedical and health informatics*, vol. 21, no. 3, pp. 803–813, 2016.

[99] B. S. Kirar and D. K. Agrawal, "Computer aided diagnosis of glaucoma using discrete and empirical wavelet transform from fundus images," *IET Image Processing*, vol. 13, no. 1, pp. 73–82, 2019.

[100] P. K. Chaudhary and R. B. Pachori, "Automatic diagnosis of glaucoma using two-dimensional fourier-bessel series expansion based empirical wavelet transform," *Biomedical Signal Processing and Control*, vol. 64, p. 102237, 2021.

[101] ——, "Denoising of biomedical images using two-dimensional fourier-bessel series expansion based empirical wavelet transform."

[102] S. Mohapatra, G. K. Pati, M. Mishra, and T. Swarnkar, "Gastrointestinal abnormality detection and classification using empirical wavelet transform and deep convolutional neural network from endoscopic images," *Ain Shams Engineering Journal*, vol. 14, no. 4, p. 101942, 2023.

[103] B. Singha Deo, M. Pal, P. K. Panigrahi, and A. Pradhan, "An ensemble deep learning model with empirical wavelet transform feature for oral cancer histopathological image classification," *medRxiv*, pp. 2022–11, 2022.

[104] P. Gaur, V. Malaviya, A. Gupta, G. Bhatia, R. B. Pachori, and D. Sharma, "Covid-19 disease identification from chest ct images using empirical wavelet transformation and transfer learning," *Biomedical Signal Processing and Control*, vol. 71, p. 103076, 2022.

[105] R. K. Patel and M. Kashyap, "Machine learning-based lung disease diagnosis from ct images using gabor features in littlewood paley empirical wavelet transform (lpewt) and lle," *Computer Methods in Biomechanics and Biomedical Engineering: Imaging & Visualization*, pp. 1–15, 2023.

[106] B. K. Swain, M. Sahani, and R. Sharma, "Empirical wavelet transform-based fast deep convolutional neural network for detection and clas-

sification of melanoma," in *Computational Intelligence in Cancer Diagnosis*. Elsevier, 2023, pp. 237–250.

[107] A. Alloqmani, Y. B. Abushark, A. I. Khan, and F. Alsolami, "Deep learning based anomaly detection in images: insights, challenges and recommendations," *International Journal of Advanced Computer Science and Applications*, vol. 12, no. 4, 2021.

[108] T. Ng, "Inbreast 2012," https://www.kaggle.com/datasets/tommyngx/inbreast2012, 2012, 25/09/2023.

[109] M. Ahmadi, M. Sabokrou, M. Fathy, R. Berangi, and E. Adeli, "Generative adversarial irregularity detection in mammography images," in *International Workshop on PRedictive Intelligence In MEdicine*. Springer, 2019, pp. 94–104.

[110] J. Zhang, Y. Xie, G. Pang, Z. Liao, J. Verjans, W. Li, Z. Sun, J. He, Y. Li, C. Shen, and Y. Xia, "Viral pneumonia screening on chest x-rays using confidence-aware anomaly detection," *IEEE Transactions on Medical Imaging*, vol. 40, no. 3, pp. 879–890, 2021.

[111] J. Amin, M. Sharif, A. Haldorai, M. Yasmin, and R. S. Nayak, "Brain tumor detection and classification using machine learning: a comprehensive survey," *Complex & intelligent systems*, pp. 1–23, 2021.

[112] K. Ouardini, H. Yang, B. Unnikrishnan, M. Romain, C. Garcin, H. Zenati, J. P. Campbell, M. F. Chiang, J. Kalpathy-Cramer, V. Chandrasekhar *et al.*, "Towards practical unsupervised anomaly detection on retinal images," in *Domain Adaptation and Representation Transfer and Medical Image Learning with Less Labels and Imperfect Data: First MICCAI Workshop, DART 2019, and First International Workshop, MIL3ID 2019, Shenzhen, Held in Conjunction with MICCAI 2019, Shenzhen, China, October 13 and 17, 2019, Proceedings 1*. Springer, 2019, pp. 225–234.

[113] K. Zhou, Y. Xiao, J. Yang, J. Cheng, W. Liu, W. Luo, Z. Gu, J. Liu, and S. Gao, "Encoding structure-texture relation with p-net for anomaly detection in retinal images," in *Computer Vision–ECCV 2020: 16th European Conference, Glasgow, UK, August 23–28, 2020, Proceedings, Part XX 16*. Springer, 2020, pp. 360–377.

[114] J. A. Jablonski, T. J. Bihl, and K. W. Bauer, "Principal component reconstruction error for hyperspectral anomaly detection," *IEEE Geoscience and Remote Sensing Letters*, vol. 12, no. 8, pp. 1725–1729, 2015.

[115] S. Tian, J. Yu, and C. Yin, "Anomaly detection using support vector machines," in *Advances in Neural Networks–ISNN 2004: International Symposium on Neural Networks, Dalian, China, August 2004, Proceedings, Part I 1*. Springer, 2004, pp. 592–597.

[116] O. Nizan and A. Tal, "k-nnn: Nearest neighbors of neighbors for anomaly detection," *arXiv preprint arXiv:2305.17695*, 2023.

[117] H. Jia and W. Liu, "Anomaly detection in images with shared autoencoders," *Frontiers in Neurorobotics*, vol. 16, p. 1046867, 2023.

[118] Y. Lu and P. Xu, "Anomaly detection for skin disease images using variational autoencoder," *arXiv preprint arXiv:1807.01349*, 2018.

[119] L. Deecke, R. Vandermeulen, L. Ruff, S. Mandt, and M. Kloft, "Image anomaly detection with generative adversarial networks," in *Machine Learning and Knowledge Discovery in Databases: European Conference, ECML PKDD 2018, Dublin, Ireland, September 10–14, 2018, Proceedings, Part I 18*. Springer, 2019, pp. 3–17.

[120] C. J. Mungall, C. Torniai, G. V. Gkoutos, S. E. Lewis, and M. A. Haendel, "Uberon, an integrative multi-species anatomy ontology," *Genome biology*, vol. 13, no. 1, p. R5, 2012.

[121] N. F. Noy, M. Crubézy, R. W. Fergerson, H. Knublauch, S. W. Tu, J. Vendetti, and M. A. Musen, "Protégé-2000: an open-source ontology-development and knowledge-acquisition environment." in *AMIA... Annual Symposium proceedings. AMIA Symposium*, vol. 2003. American Medical Informatics Association, 2003, pp. 953–953.

[122] S. Falconer, "Ontograf protege plugin," *Place: Available at: http://protegewiki. stanford. edu/wiki/OntoGraf [Accessed: 21/03/2014]*, 2010.

[123] E. Loper and S. Bird, "Nltk: the natural language toolkit," *arXiv preprint cs/0205028*, 2002.

[124] A. S. Boerlage, A. Ashby, J. I. Eze, G. J. Gunn, and A. Reeves, "Field evaluation of diagnostic sensitivity (dse) and specificity (dsp) of common tests for amoebic gill disease (agd) and complex gill disease (cgd) in cultured Atlantic salmon (Salmo salar) in Scotland using bayesian latent class models," *Preventive Veterinary Medicine*, vol. 204, p. 105654, 2022.

[125] A. Vahadane, T. Peng, A. Sethi, S. Albarqouni, L. Wang, M. Baust, K. Steiger, A. M. Schlitter, I. Esposito, and N. Navab, "Structure-preserving color normalization and sparse stain separation for histological images," *IEEE transactions on medical imaging*, vol. 35, no. 8, pp. 1962–1971, 2016.

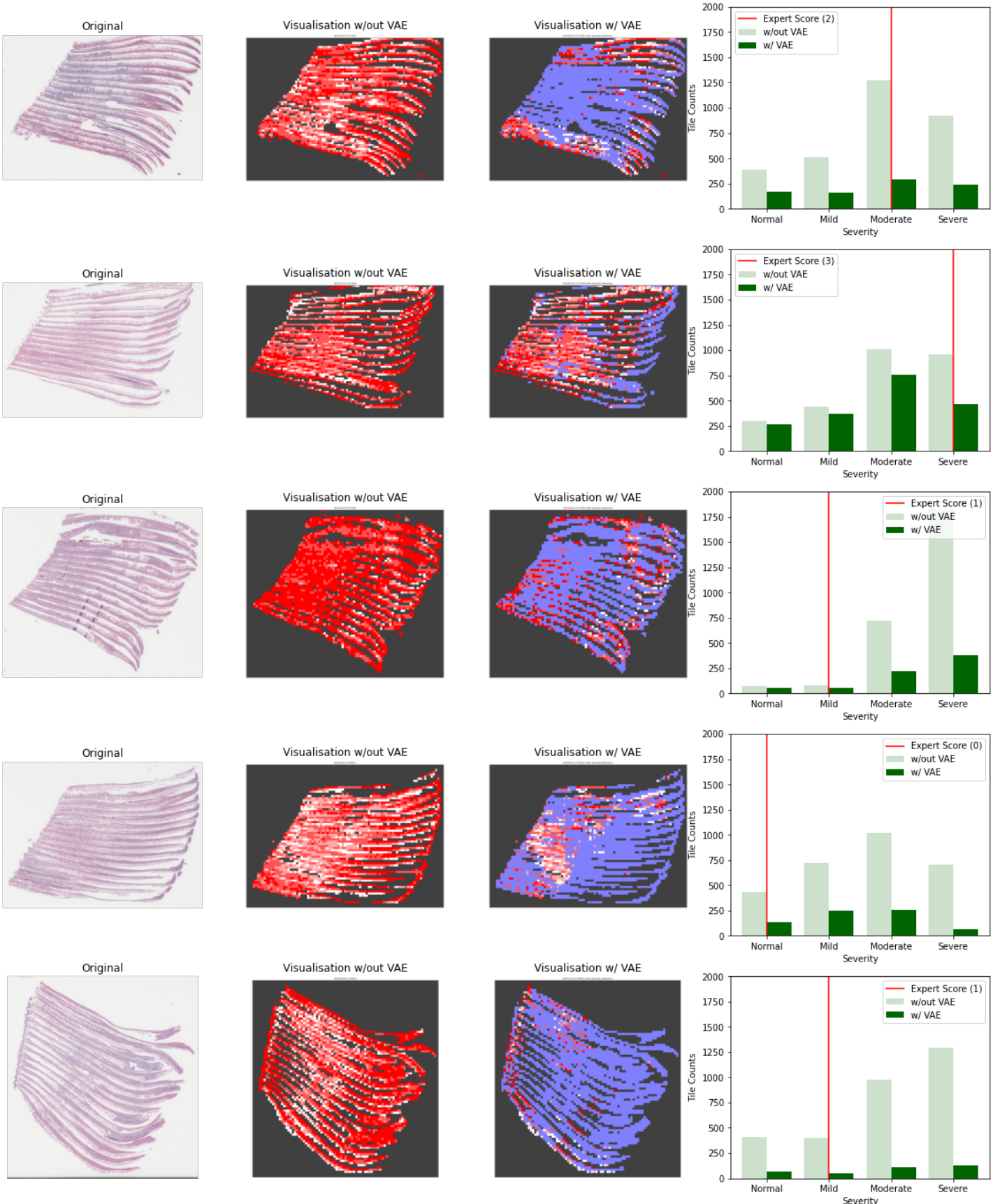
[126] D. L. Ruderman, T. W. Cronin, and C.-C. Chiao, "Statistics of cone responses to natural images: implications for visual coding," *JOSA A*, vol. 15, no. 8, pp. 2036–2045, 1998.

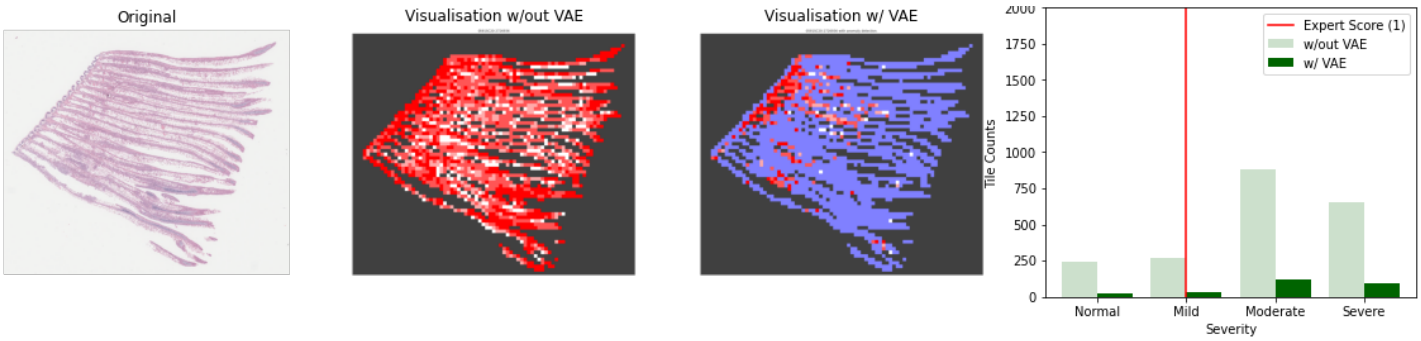
[127] S. Roy, S. Panda, and M. Jangid, "Modified reinhard algorithm for color normalization of colorectal cancer histopathology images," in *2021 29th European Signal Processing Conference (EUSIPCO)*. IEEE, 2021, pp. 1231–1235.

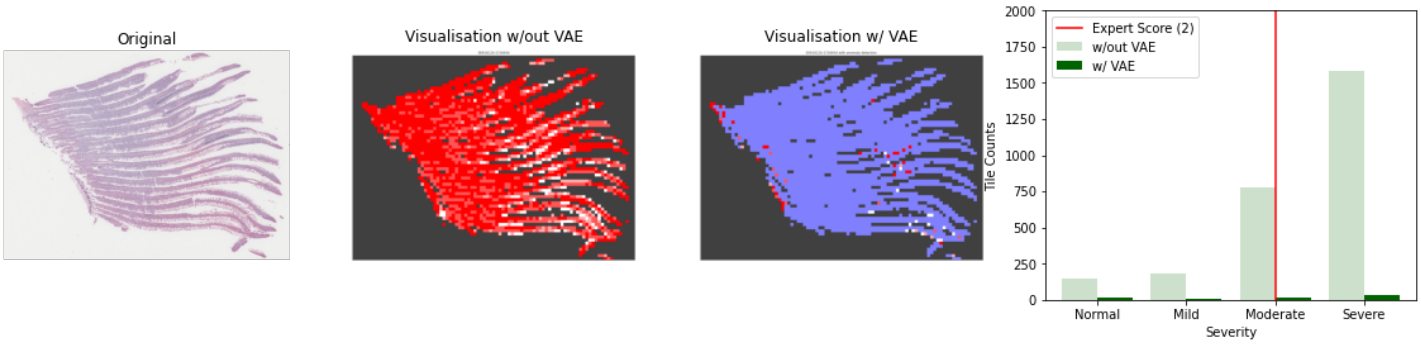
[128] K. He, X. Zhang, S. Ren, and J. Sun, "Deep residual learning for image recognition," in *Proceedings of the IEEE conference on computer vision and pattern recognition*, 2016, pp. 770–778.

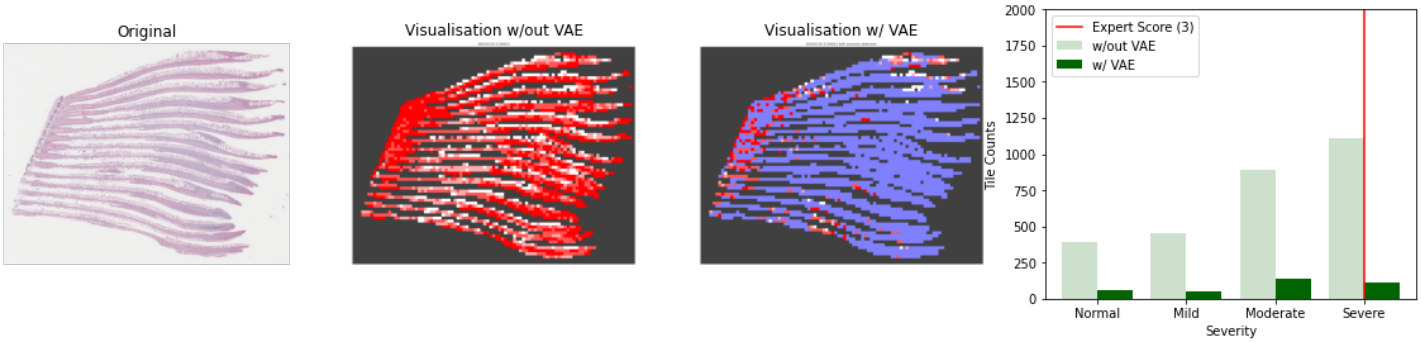
[129] J. Deng, W. Dong, R. Socher, L.-J. Li, K. Li, and L. Fei-Fei, "Imagenet: A large-scale hierarchical image database," in *2009 IEEE conference on computer vision and pattern recognition*. Ieee, 2009, pp. 248–255.

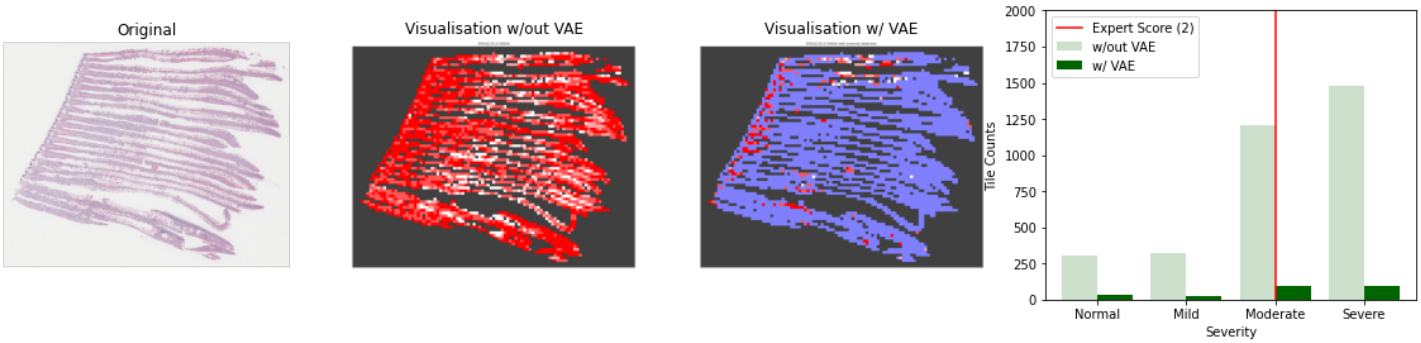
Appendix 1 - Empirical Wavelet Transform Visualisations

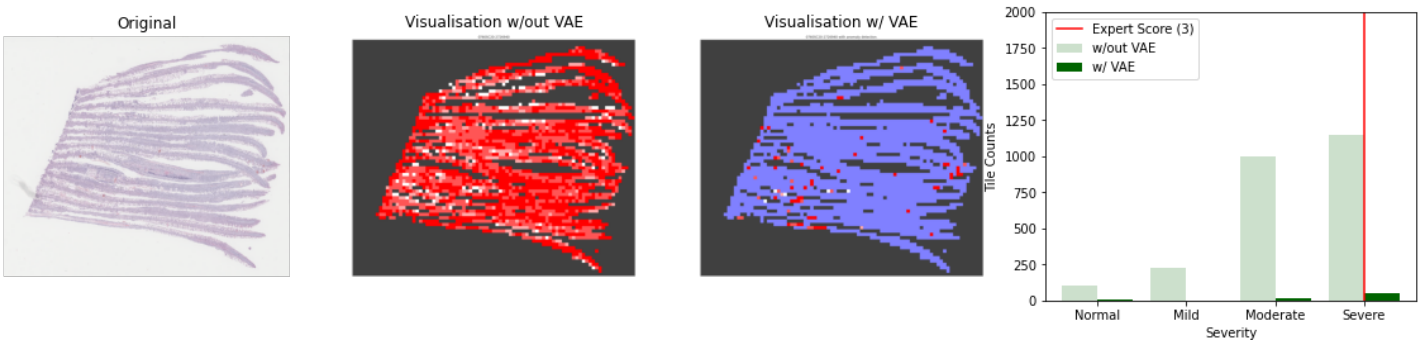



Original
Visualisation w/out VAE
Visualisation w/ VAE
Expert Score (1)
w/out VAE
w/ VAE
Tile Counts
Normal
Mild
Moderate
Severe
Severity

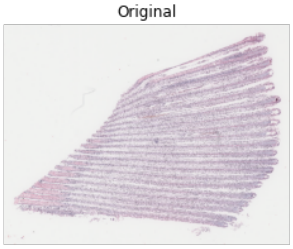

Original
Visualisation w/out VAE
Visualisation w/ VAE
Expert Score (2)
w/out VAE
w/ VAE
Tile Counts
Normal
Mild
Moderate
Severe
Severity


Original
Visualisation w/out VAE
Visualisation w/ VAE
Expert Score (3)
w/out VAE
w/ VAE
Tile Counts
Normal
Mild
Moderate
Severe
Severity

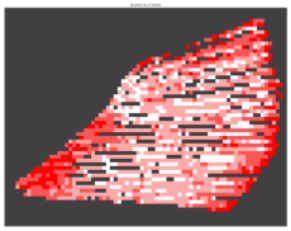

Original
Visualisation w/out VAE
Visualisation w/ VAE
Expert Score (2)
w/out VAE
w/ VAE
Tile Counts
Normal
Mild
Moderate
Severe
Severity


Original
Visualisation w/out VAE
Visualisation w/ VAE
Expert Score (3)
w/out VAE
w/ VAE
Tile Counts
Normal
Mild
Moderate
Severe
Severity

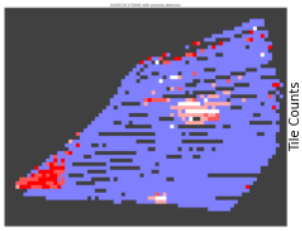
Original

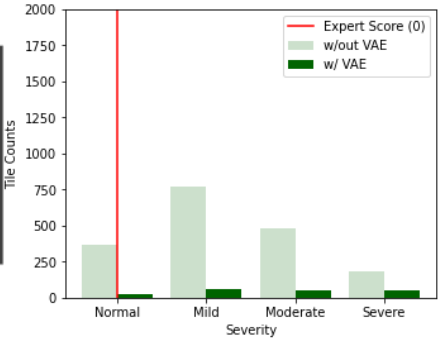


Visualisation w/out VAE

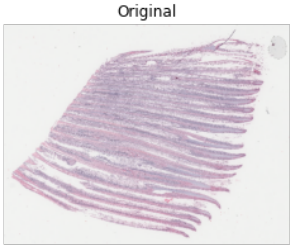


Visualisation w/ VAE

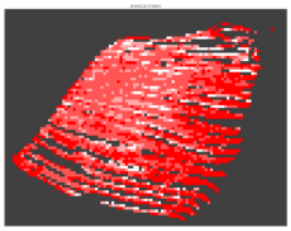




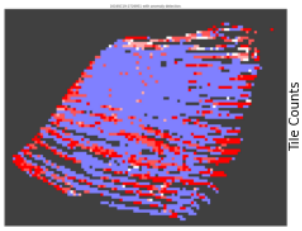
Original

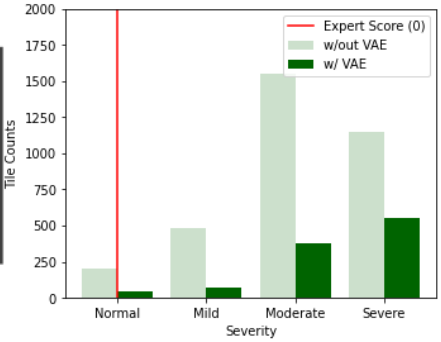


Visualisation w/out VAE

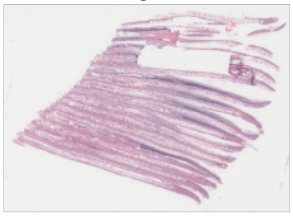


Visualisation w/ VAE

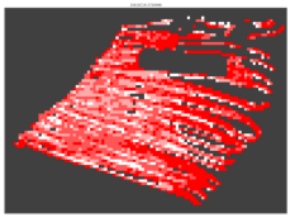




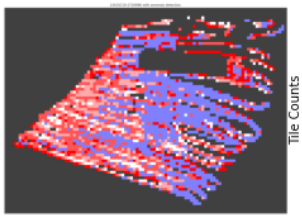
Original

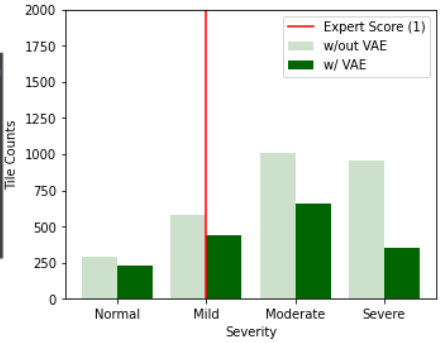


Visualisation w/out VAE

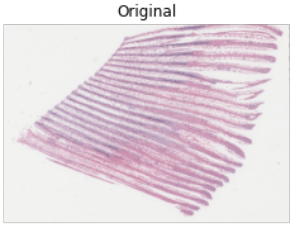


Visualisation w/ VAE

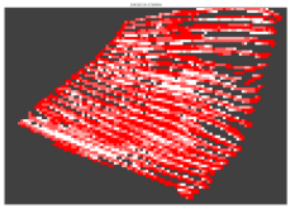




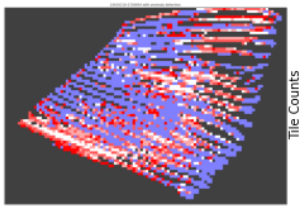
Original

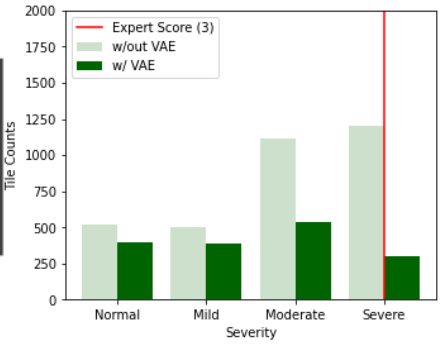


Visualisation w/out VAE

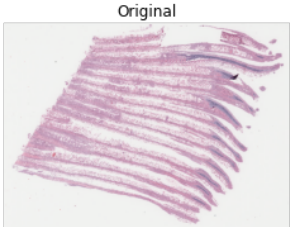


Visualisation w/ VAE

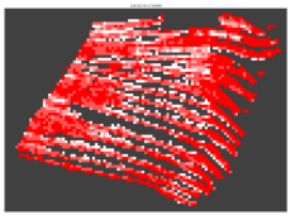




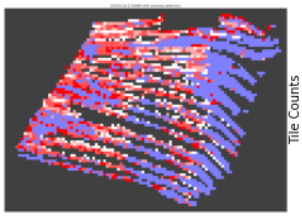
Original

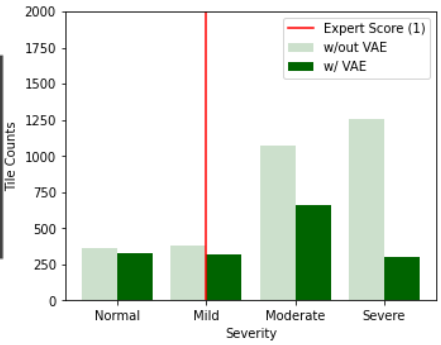


Visualisation w/out VAE

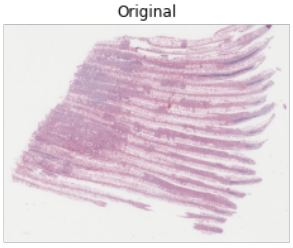


Visualisation w/ VAE

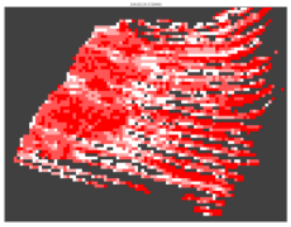




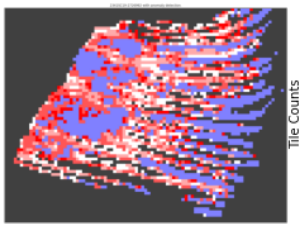
Original

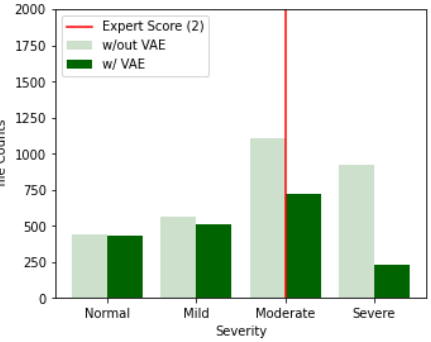


Visualisation w/out VAE

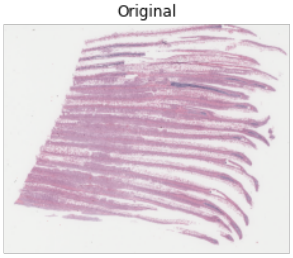


Visualisation w/ VAE

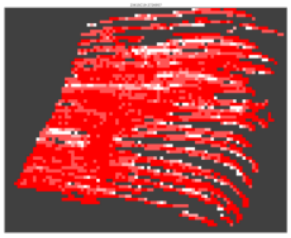




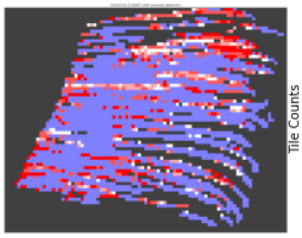
Original

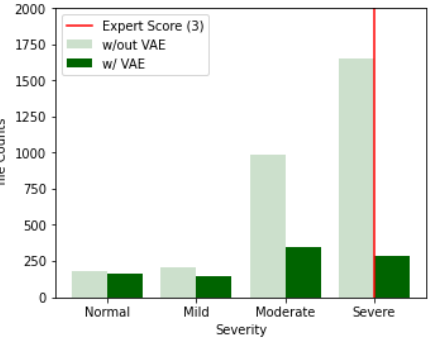


Visualisation w/out VAE

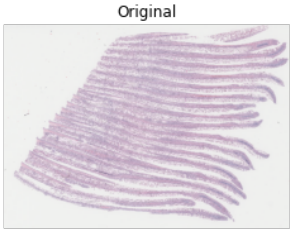


Visualisation w/ VAE

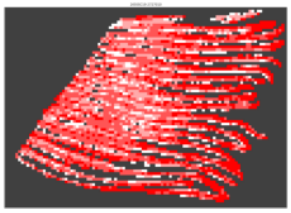




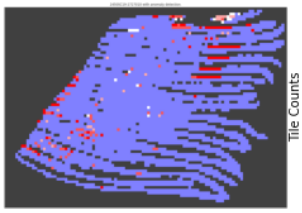
Original

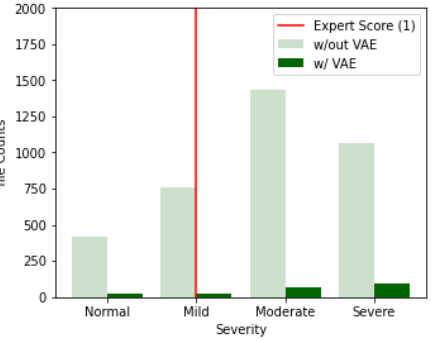


Visualisation w/out VAE

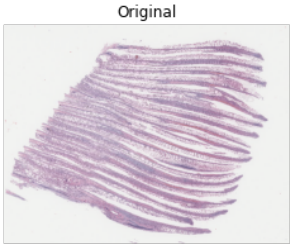


Visualisation w/ VAE

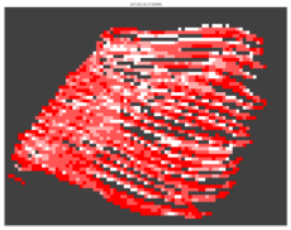




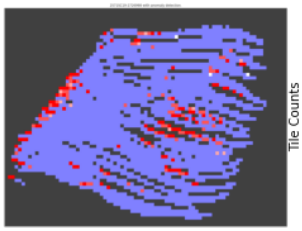
Original

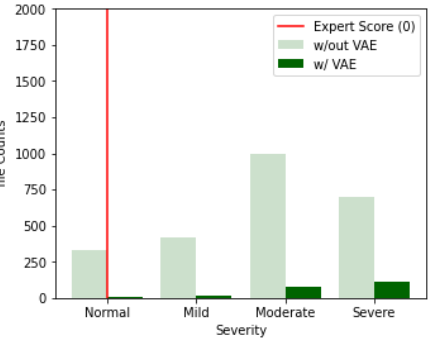


Visualisation w/out VAE

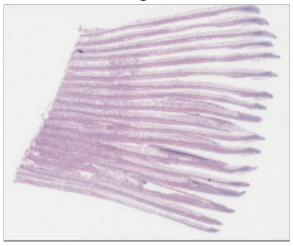


Visualisation w/ VAE

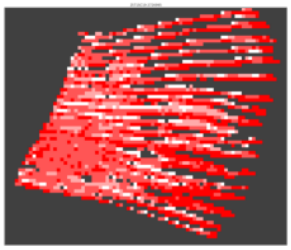




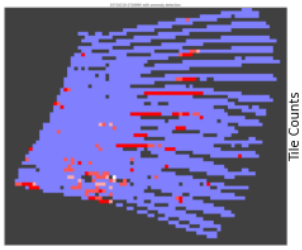
Original

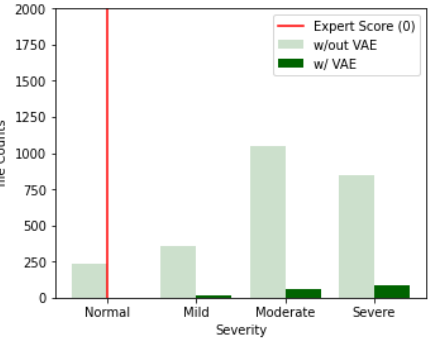


Visualisation w/out VAE

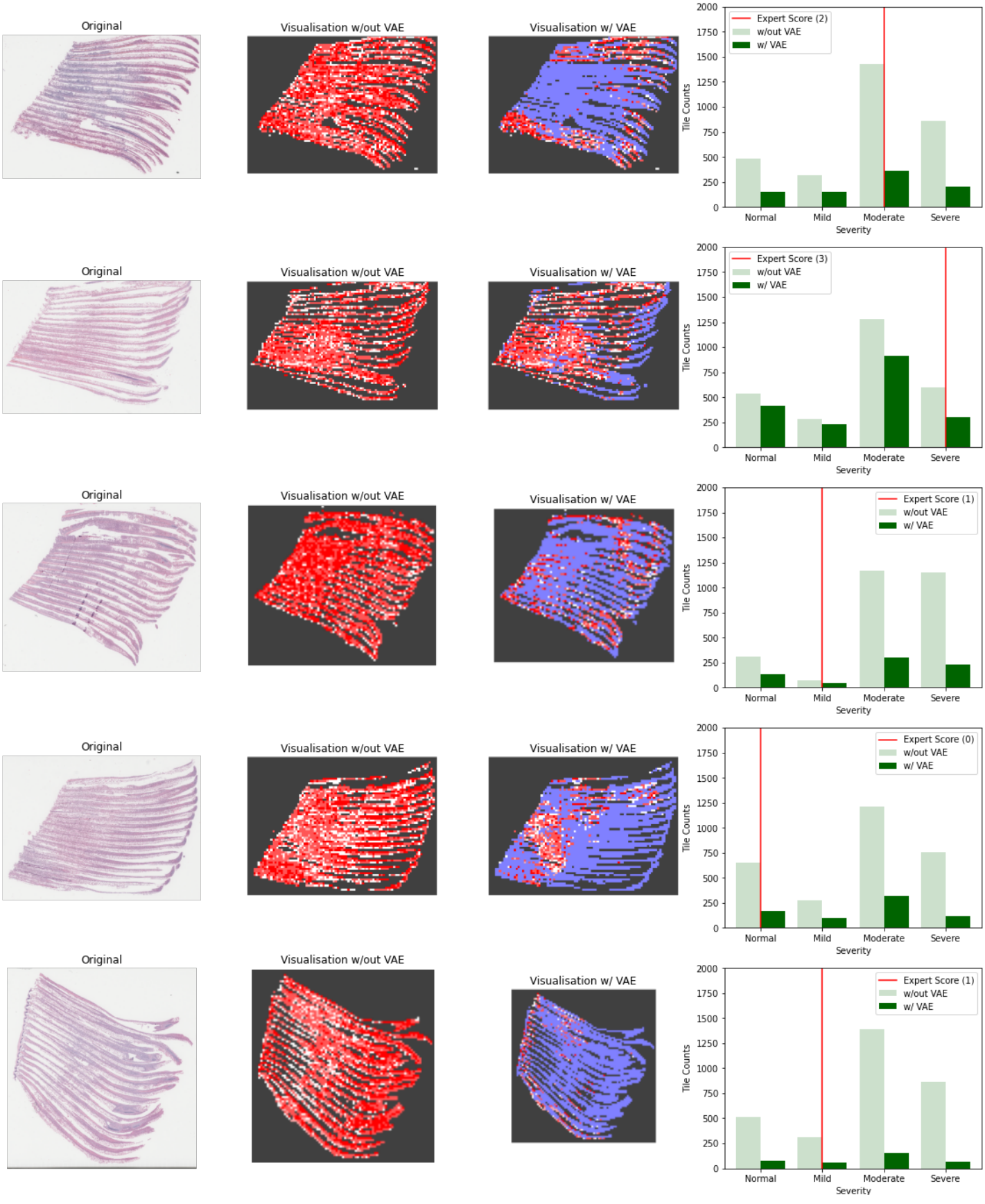


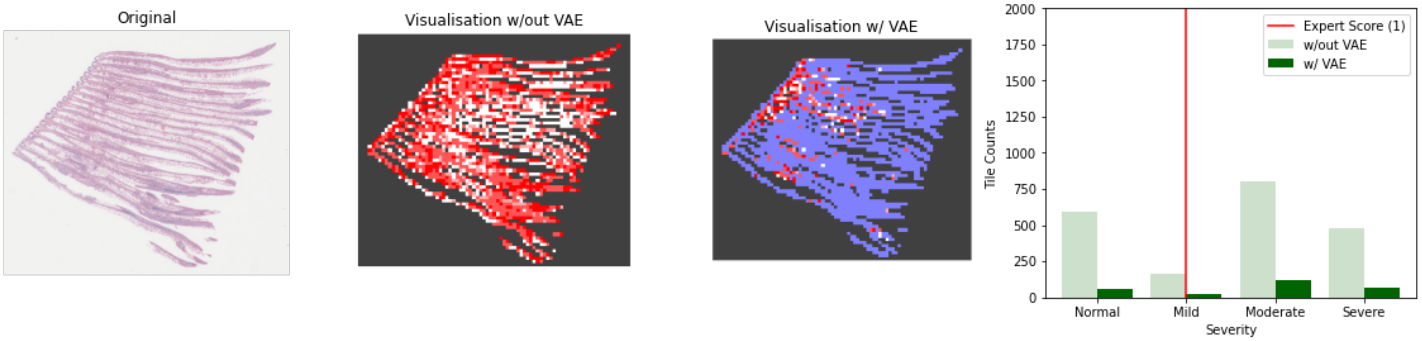
Visualisation w/ VAE

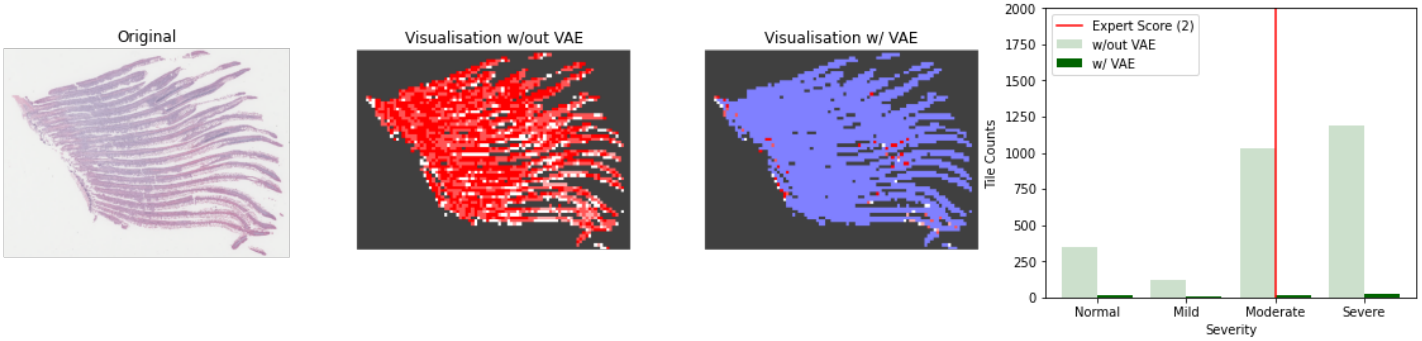


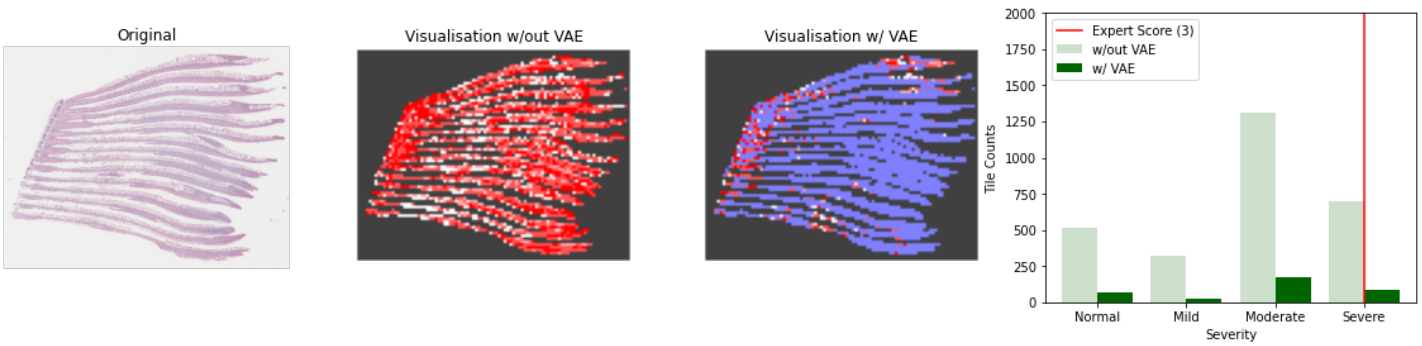


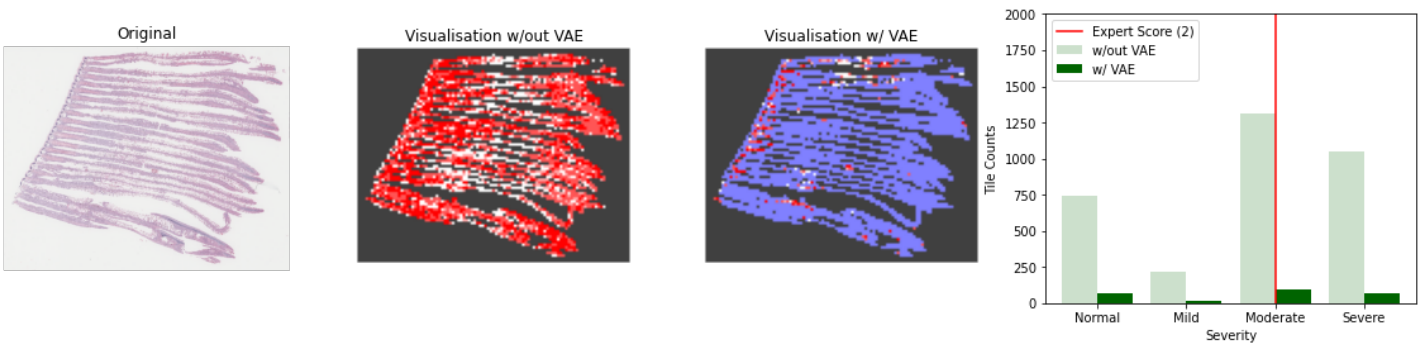
Appendix 2 - ResNet18 Visualisations

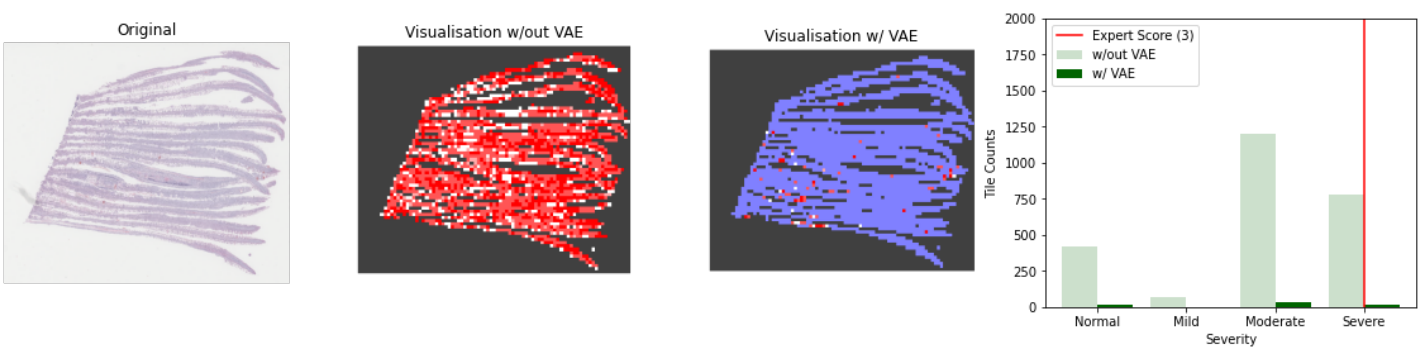


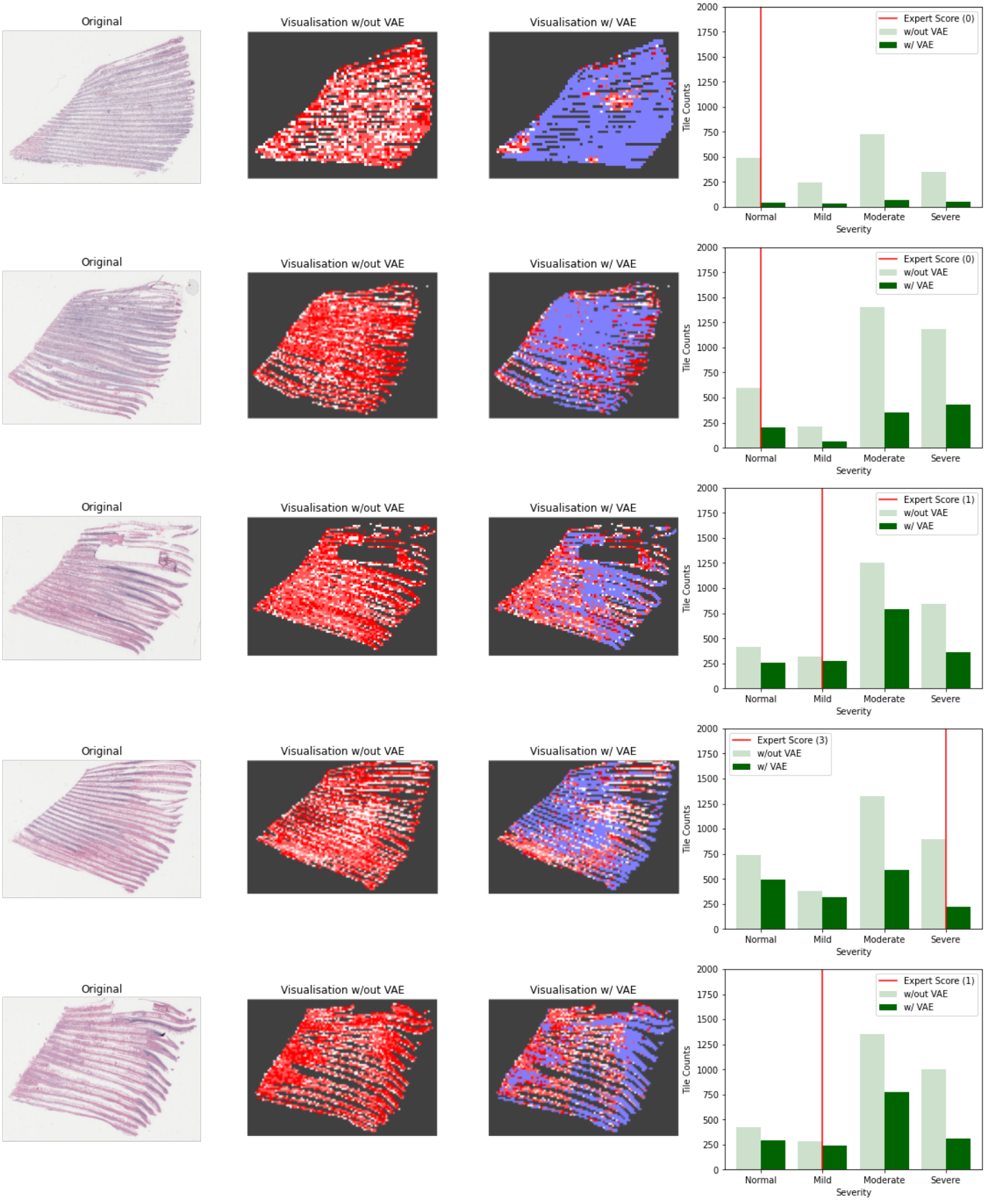

Original
Visualisation w/out VAE
Visualisation w/ VAE
Expert Score (1)
w/out VAE
w/ VAE
Tile Counts
Normal
Mild
Moderate
Severe
Severity

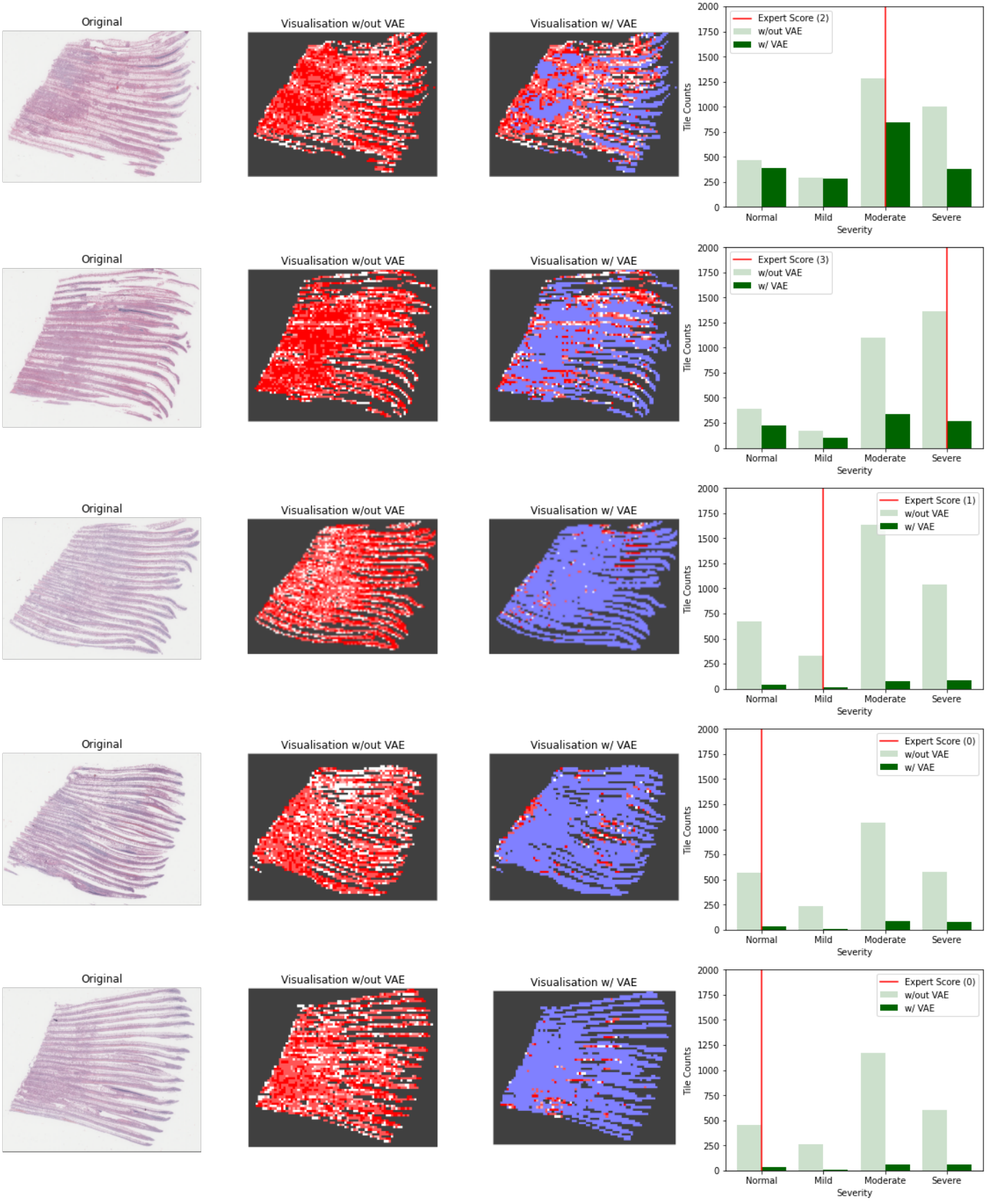

Original
Visualisation w/out VAE
Visualisation w/ VAE
Expert Score (2)
w/out VAE
w/ VAE
Tile Counts
Normal
Mild
Moderate
Severe
Severity


Original
Visualisation w/out VAE
Visualisation w/ VAE
Expert Score (3)
w/out VAE
w/ VAE
Tile Counts
Normal
Mild
Moderate
Severe
Severity


Original
Visualisation w/out VAE
Visualisation w/ VAE
Expert Score (2)
w/out VAE
w/ VAE
Tile Counts
Normal
Mild
Moderate
Severe
Severity


Original
Visualisation w/out VAE
Visualisation w/ VAE
Expert Score (3)
w/out VAE
w/ VAE
Tile Counts
Normal
Mild
Moderate
Severe
Severity


Original
Visualisation w/out VAE
Visualisation w/ VAE
Expert Score (0)
w/out VAE
w/ VAE
Tile Counts
Normal
Mild
Moderate
Severe
Severity
Original
Visualisation w/out VAE
Visualisation w/ VAE
Expert Score (0)
w/out VAE
w/ VAE
Tile Counts
Normal
Mild
Moderate
Severe
Severity
Original
Visualisation w/out VAE
Visualisation w/ VAE
Expert Score (1)
w/out VAE
w/ VAE
Tile Counts
Normal
Mild
Moderate
Severe
Severity
Original
Visualisation w/out VAE
Visualisation w/ VAE
Expert Score (3)
w/out VAE
w/ VAE
Tile Counts
Normal
Mild
Moderate
Severe
Severity
Original
Visualisation w/out VAE
Visualisation w/ VAE
Expert Score (1)
w/out VAE
w/ VAE
Tile Counts
Normal
Mild
Moderate
Severe
Severity


Original
Visualisation w/out VAE
Visualisation w/ VAE
Expert Score (2)
w/out VAE
w/ VAE
Tile Counts
Severity
Normal
Mild
Moderate
Severe
Original
Visualisation w/out VAE
Visualisation w/ VAE
Expert Score (3)
w/out VAE
w/ VAE
Tile Counts
Severity
Normal
Mild
Moderate
Severe
Original
Visualisation w/out VAE
Visualisation w/ VAE
Expert Score (1)
w/out VAE
w/ VAE
Tile Counts
Severity
Normal
Mild
Moderate
Severe
Original
Visualisation w/out VAE
Visualisation w/ VAE
Expert Score (0)
w/out VAE
w/ VAE
Tile Counts
Severity
Normal
Mild
Moderate
Severe
Original
Visualisation w/out VAE
Visualisation w/ VAE
Expert Score (0)
w/out VAE
w/ VAE
Tile Counts
Severity
Normal
Mild
Moderate
Severe

Appendix 3 - Ir-Man: An Information Retrieval Framework for Marine Animal Necropsy Analysis

Ir-Man: An Information Retrieval Framework for Marine Animal Necropsy Analysis

Alexander F. B. Carmichael*
Computing Science and Mathematics
University of Stirling
Stirling, UK
a.f.carmichael@stir.ac.uk

Deepayan Bhowmik
Computing Science and Mathematics
University of Stirling
Stirling, UK
deepayan.bhowmik@stir.ac.uk

Johanna Baily
Institute of Aquaculture
University of Stirling
Stirling, UK
j.l.baily@stir.ac.uk

Andrew Brownlow
Scottish Marine Animal Stranding Scheme
Scotland's Rural College (SRUC)
Inverness, UK
andrew.brownlow@sruc.ac.uk

George J. Gunn
Epidemiology Research Unit
Scotland's Rural College (SRUC)
Inverness, UK
george.gunn@sruc.ac.uk

Aaron Reeves
Epidemiology Research Unit
Scotland's Rural College (SRUC)
Inverness, UK
aaron.reeves@sruc.ac.uk

ABSTRACT

This paper proposes Ir-Man (Information Retrieval for Marine Animal Necropsies), a framework for retrieving discrete information from marine mammal post-mortem reports for statistical analysis. When a marine mammal is reported dead after stranding in Scotland, the carcass is examined by the Scottish Marine Animal Strandings Scheme (SMASS) to establish the circumstances of the animal's death. This involves the creation of a 'post-mortem' (or necropsy) report, which systematically describes the body. These semi-structured reports record lesions (damage or abnormalities to anatomical regions) as well as other observations. Observations embedded within these texts are used to determine cause of death. While a cause of death is recorded separately, many other descriptions may be of pathological and epidemiological significance when aggregated and analysed collectively. As manual extraction of these descriptions is costly, time consuming and at times erroneous, there is a need for an automated information retrieval mechanism which is a non-trivial task given the wide variety of possible descriptions, pathologies and species. The Ir-Man framework consists of a new ontology, a lexicon of observations and anatomical terms and an entity relation engine for information retrieval and statistics generation from a pool of necropsy reports. We demonstrate the effectiveness of our framework by creating a rule-based binary classifier for identifying bottlenose dolphin attacks (BDA) in harbour porpoise gross pathology reports and achieved an accuracy of 83.4%.

CCS CONCEPTS

• **Applied computing** → **Bioinformatics**; • **Information systems** → **Ontologies**; **Information extraction**.

KEYWORDS

Information retrieval, marine animal, necropsy analysis, ontology.

ACM Reference Format:
Alexander F. B. Carmichael, Deepayan Bhowmik, Johanna Baily, Andrew Brownlow, George J. Gunn, and Aaron Reeves. 2020. Ir-Man: An Information Retrieval Framework for Marine Animal Necropsy Analysis. In *Proceedings of the 11th ACM International Conference on Bioinformatics, Computational Biology and Health Informatics (BCB '20), September 21–24, 2020, Virtual Event, USA.* ACM, New York, NY, USA, 9 pages. https://doi.org/10.1145/3388440.3412417

1 INTRODUCTION

Monitoring and surveillance of wildlife is fundamental for the development of understanding of the factors which impact the well-being of populations, species and ecosystems. These activities are especially difficult when applied to the marine mammal domain, as direct observation of living animals in their environments is often impractical. Observation of dead animals, when they become accessible, provides a critical source of data for our knowledge of these populations, and information gathered from such events is particularly important. When a cetacean or pinniped becomes stranded and dies, and when its carcass is examined by trained investigators, the resulting post-mortem (PM) report provides a snapshot of the animal's condition. Collectively, such data provides a unique insight into the general welfare of marine mammal populations and may reveal problems facing the species' environment as a whole.

Williams *et al.* [23, 24], for example have extensively monitored the levels of toxic polychlorinated biphenyls (PCBs) in harbour porpoises. PCB levels are directly affected by human pollution due to the compound's use in some manufactured goods. Similarly, Nelms *et al.* [19] have analysed the presence of microplastics found in stranded cetacea using PM examinations. These examples show that PM examinations can be used to observe the human impact on marine ecology.

*Also with Epidemiology Research Unit, Scotland's Rural College (SRUC).

BCB '20, September 21–24, 2020, Virtual Event, USA

ACM ISBN 978-1-4503-7964-9/20/09…$15.00
https://doi.org/10.1145/3388440.3412417

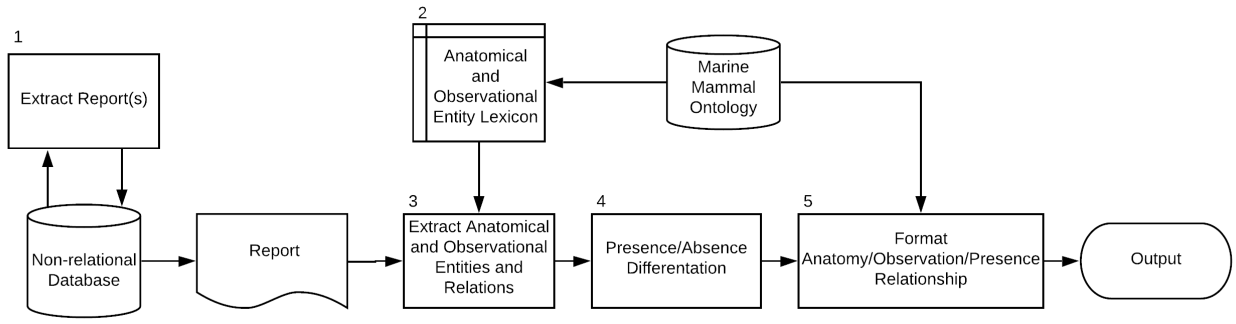


Figure 1: Overview of the Ir-Man framework along with its information retrieval flow diagram.

PM reports are generated in a semi-structured format, with information embedded across multiple free text sections which makes the retrieval of pathological findings a non-trivial problem. Furthermore, there are many cases where multiple indicators are described which relate to multiple distinct pathologies. The ability to effectively confirm or rule out the presence of a pathology based on descriptions of abnormalities would allow for clearer understanding of the problems facing marine mammals. Information retrieval approaches have been applied extensively in human pathology [3, 8, 11, 21, 25, 26], and other animal pathologies [2, 7, 13], however, no work currently exists for marine mammal pathology free text.

To address such gaps we propose Ir-Man (Information Retrieval for Marine Animal Necropsies), a new framework for retrieving discrete information from marine mammal post-mortem (necropsy) reports for statistical analysis. We infer that the hierarchical nature of descriptive terms used in marine mammal PM reports can be represented within an ontology, in which a class of terms represent a specific observation, and subclasses represent distinct varieties of observations. Our approach involves the extraction of a number of key pieces of information: the term used to describe particular observations; whether the context indicates presence or absence; and the anatomical region to which the observation relates. Once extracted, these fields can be used to create a deterministic classifier based on the presence or absence of either general pathological indicators or indicators of pathologies in specific anatomical regions. We also extract a reference that can be used to link and aggregate information across different report types. The overview of Ir-Man framework is depicted in Figure 1.

In building the framework, this paper considers the marine mammal PM reports, particularly the gross pathology reports and describes various framework components for extraction of gross pathology findings. In evaluating the effectiveness of Ir-Man, we experimented with an exemplar use case of identifying bottlenose dolphin attacks on harbour porpoises and reported our findings. Our main contributions are:

- We propose a new framework for information retrieval for marine mammal necropsy analysis using an ontology driven entity relation approach.
- We design three ontologies which contain terms relevant to cetacean gross pathology reports that are based on observations, anatomy, and pathology, respectively.
- We develop a lexicon based entity-relationship engine that can record the presence or absence of observations which can confirm or rule out pathologies.
- We measure the effectiveness of our retrieval approach by creating and evaluating a deterministic classifier for cases of bottlenose dolphin attacks (BDAs) on harbour porpoises.

2 RELATED WORK

Our framework encompasses several information retrieval methods and fields. While current literature covers a relatively large number of information retrieval approaches in the biomedical domain (particularly on human health), only a handful of attempts were made in the veterinary domain. We critically analysed both domains separately in order to position our proposed framework appropriately.

2.1 Information Retrieval in Biomedical Domain

Related approaches for information extraction in the biomedical domain are numerous. Chapman *et al.* [3] proposed NegEx, a tool for determining the presence or absence of clinical findings in discharge summaries. Their approach was used to analyse $76,049$ screening and $17,656$ diagnostic mammography reports. Even though this approach extracts conclusions, rather than observations – which are the focus of our work - the applications are similar. More recently Gao *et al.* [8] extracted a number of features from mammographic reports: mass, calcification, asymmetry and architectural distortion.

Friedlin *et al.* [6] developed Medical Exploratory Data Analysis over Text (MEDAT), a text analytics system for medical domains and demonstrated their system on radiology reports. Comelli *et al.* [4] also applied text mining to radiology reports. They leveraged the entity relationships represented in their radiology ontology, to extract relevant medical terms from mammographic reports. While their approach was exhaustive within the mammographic domain, the texts were in Italian and the application domain is far more specific than that of marine mammal gross pathology.

Gong *et al.* [11] developed a biomedical information retrieval approach for terminologies related to breast cancer. Their approach

involves entity extraction, entity relationship identification, and visualisation. Entities are extracted based on conditional random fields while entity relationships were extracted using co-occurrence statistics. Sudeshna *et al.* [21] aimed to identify symptoms and treatments of heart disease using a machine learning based approach. Based on suggested identified symptoms, texts would be classified into treatments. Zhao *et al.* [26] created CausalTriad, an approach toward the discovery of pseudo causal relationships between entities. They evaluated their approach on HealthBoards message board data and Traditional Chinese Medicine data. Yang *et al.* [25] used an ontology-based text mining approach for the extraction of data from Chinese EMRs. This work focused on the mining of stroke cases. Gero and Ho [9] proposed NamedKeys, a keyphrase extraction approach which they evaluated on PubMed abstracts. They also describe a benchmark dataset for biomedical keyphrase extraction.

While a variety of different clinical text types have been the subject of such research, there is also a wide body of research into automated biomedical literature reviews. Navathe [18] used UMLS (Unified Medical Language System) [1] and a gene ontology to represent biomedical concepts, and an SVM to classify literature from the Centre of Disease Control (CDC) based on relevant keywords. Mala *et al.* [15] researched the use of ontology in semantic medical text mining with WordNet. Gong *et al.* [10] used a dictionary-based approach to extract biomedical concepts from literature. This was done using an algorithm called the Variable-step Window Identification Algorithm (VWIA), matched terms to biomedical entities using POS tagging and organisation based on phrasing. Their technique was applied to 10 Medline abstracts and produced promising results. Mate *et al.* [16] focused on creating a process of extraction, transformation and loading (ETL) of electronic medical records.

Although not used in our approach, it should be noted that emerging *deep learning* has become popular in the biomedical domain with neural network based methods being used to enhance text mining techniques [12, 22].

2.2 Information Retrieval in Veterinary Domain

All of the reports listed above applied information retrieval techniques to biomedical text pertaining to humans. In the veterinary domain, Bollig *et al.* [2] used a machine learning based approaches for extraction of different pathologies from free text. Furrer *et al.* [7] built a text mining tool for veterinary surveillance by linking terms identified in necropsies to existing ontologies. Küker *et al.* [13] later used this tool to analyse pig and cattle necropsies and found that free text necropsy reports are a valuable resource for animal health surveillance.

At present no work exists on information retrieval from marine mammal necropsy reports. Given the importance of PMs in furthering understanding of marine mammals and marine ecology more generally, an information retrieval framework that can aggregate observations for statistical and epidemiological analysis would be especially useful.

3 THE FRAMEWORK

In developing the proposed Ir-Man framework, we consider a number of steps that are involved in the extraction of observations from marine mammal necropsies. Firstly, free text is pulled from necropsy documents, and then individual reports sections (*i.e.*, the gross pathology report section and if applicable, the histopathology and bacteriology report sections) are extracted. Text is divided into sentence tokens before individual words are tagged based on part-of-speech. Entities and the relationships between them are then grouped using a feature-based grammar. Each entity is then checked against our lexicon of anatomical, pathological and observational terms which is generated using our ontology. Presence or absence of a described feature is then established by checking for negation. It is fundamental to record negative occurrences of identifiers (absences) as well as positive occurrences, (presence) as both can be leveraged in a deterministic classification system. The overall retrieval process is outlined in Figure 1 and the pseudo-code in Algorithm 1 along with the description of each framework components below.

Algorithm 1: Information retrieval pipeline. Output of the entity-relationship extraction engine is used to identify observations, attributed anatomies and detect negation.

```
Result: relationships
sentences = sentenceTokenisation(text);
observations;
while not at end of sentences do
    RELChunkedSentence = preprocess(sentence);
    identifyNamedEntities(RELChunkedSentence);
    while not at end of sentences do
        if No Observational Entities then
            break to next relationship;
        end
        if Observational Entity AND No Anatomical Entity
         then
            observations <- 'unattributed' observation;
            break to next relationship
        end
        if Observational Entity and Anatomical Entity then
            observations <- anatomy, observation;
        end
    end
end
while not at end of observations do
    negatedObservation <- mark_negated(observation) ;
    if observation == negatedObservation then
        presence <- True;
    else
        presence <- False;
    end
end
```

3.1 Data Set: Marine Mammal Stranding Reports

The data used in this project was generated using PM reports of cetacea produced by the Scottish Marine Animal Stranding Scheme (SMASS) between 2012 and 2019. When generating these reports, the pathologist records features of the carcass, including condition, morphology, pathological lesions and observations. Information relating to the body's condition usually refers to the level of autolysis

Body condition: Fat

External examination
Body orifices: NAD
Ectoparasites: NAD – None seen
Fins and flukes: NAD – Intact, no rake marks

Integument
Epidermis: Rake marks over left flank/tailstock. Severe scavenger damage at right side of head
Blubber: NAD – Good layer, not jaundiced
Subcutaneous tissue: Bruising over lateral spinous processes and right side head region
Mammary glands: NE

Figure 2: A gross pathology extract from a harbour porpoise necropsy report.

or physical damage to the remains. Morphometric measurements, such as blubber thickness and body length, are also taken. These features all help inform the pathologist of a probable cause of death. PM reports include gross pathology reports, which describe, in detail, the characteristics of the body as a whole, and those of specific anatomies. The final PM report then contains a number of sections: basic information (sex, date, location etc.); morphometric data; a gross pathology; bacteriology and histopathology reports, where applicable; and a conclusion which includes comments, cause of death, and an indicator of confidence in diagnosis.

The material analysed for this paper consists of 193 gross pathology reports on harbour porpoises (*Phocoena phocoena*). This species was chosen for a number of reasons: the relatively high number of reports produced by SMASS on harbour porpoises; easy future integration of other cetacean species due to transferability of the harbour porpoise anatomy; and the prevalence of BDAs listed as the cause of death, which allows us to establish the suitability of our framework for detecting exhibited pathologies. Bottlenose dolphins are known to violently attack harbour porpoises, usually leaving parallel incisions which are referred to as 'rake marks'. It is these rake marks which are used as a primary indicator of a BDA, and as such, the use of the term is relatively consistent, making it a good candidate for evaluating the effectiveness of our approach.

While the language used in these gross pathology reports is specialised, there is some structure to the reports which can be leveraged. An anatomical region of interest will often be used as a heading followed by a free text description. This can be seen in Figure 2. Acronyms such as NAD (no abnormalities detected) and NE (not examined) are also important and distinct. One can rule out some pathological conditions when no abnormality is detected, but not when a region has not been examined.

3.2 Gross Pathology Report Extraction

The SMASS post-mortem reports were stored in Microsoft Word Open XML Format (DOCX) files. We parsed documents and stored fields in a non-relational MongoDB[1] database. Where applicable, specific text fields were extracted by searching for field names which where indicative of a field's presence. An example would the species field, where we used the string "SPECIES:" as the field indicator and the string following it in the line as the field to be extracted (*e.g.*, "*delphinus delphis*"). When a field was left blank, no value was stored in the database. We normalised fields by grouping synonymous terms. For example, the case of the species field this involved pairing the scientifc names (*e.g.*, "*delphinus delphis*") with their corresponding common names (*e.g.*, "short-beaked common dolphin"). Free text sections such as the gross pathology reports were obtained by identifying relevant section headers and extracting the text between them. When the space between section headers consisted only of white-space or short strings such as "Not examined", the section was not extracted. All extracted fields and sections were stored in a local MongoDB database.

3.3 Ontology Development

The framework uses our bespoke ontologies to organise terms, and to provide context that would otherwise be unavailable. While multi-species ontologies such as Uberon [17] do exist, it was decided that a smaller more manageable ontology would be more appropriate for this task. We identified three main branches of relevant terminologies for our purposes. The first is a representation of anatomy, where classes represent different anatomical regions. The second is the pathology ontology which was used to record different conditions which can be represented in PM reports. The third is the observation ontology, which groups terms into classes and sub-classes where children represent an extra degree of specification that may not apply to all within the parent class. For all classes a representative label is stored in the "rdfs:label" annotation, and manually generated synonymous terms are stored in our own "synonym" annotation. Ontologies were developed using Protégé [20] (shown in Figure 3) and stored in RDF/XML format.

3.3.1 Observation Ontology. The observation branch of the ontology makes use of the semantic relationships between terms. When terms are very similar semantically, but one gives a greater degree of specification, the more specific term is considered a child of the other. For example, reports may specify that "fluid" or "brown fluid" is present. Not all fluid is brown fluid, so a parent-child relationship is created between the terms. This allows for distinctions between different types of fluids and their descriptions such as mucoid, protein-rich or amniotic fluids. The ontology was populated manually by producing lists of terms from the reports based frequent unigrams, bigrams, and trigrams, as well as collocations using pointwise mutual information (PMI). Previously established anatomical and disease related terminologies were filtered to accelerate the process. The structure of the observational ontology is shown in Figure 4.

3.3.2 Pathology Ontology. The pathology ontology (shown in Figure 5) is used to represent different conditions and the semantic relationships between them. This was initially created using the diseases or conditions listed as a cause of death within the SMASS database. These were also mined from the reports using known target strings

[1] https://www.mongodb.com/

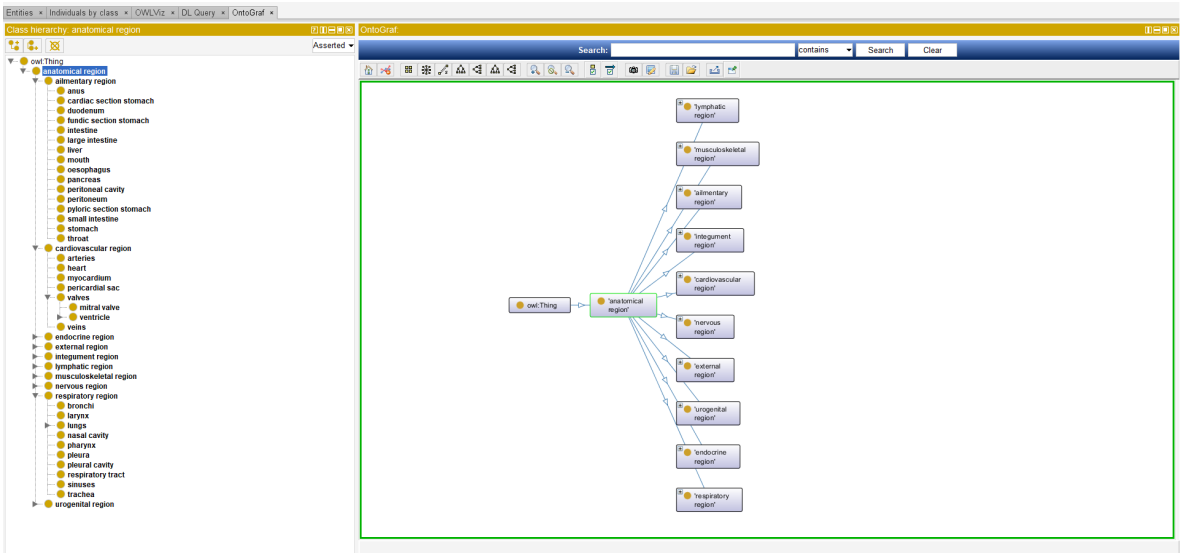


Figure 3: Screenshot of Protégé IDE for ontology development. The OntoGraf plugin [5] was used for ontology visualisation.

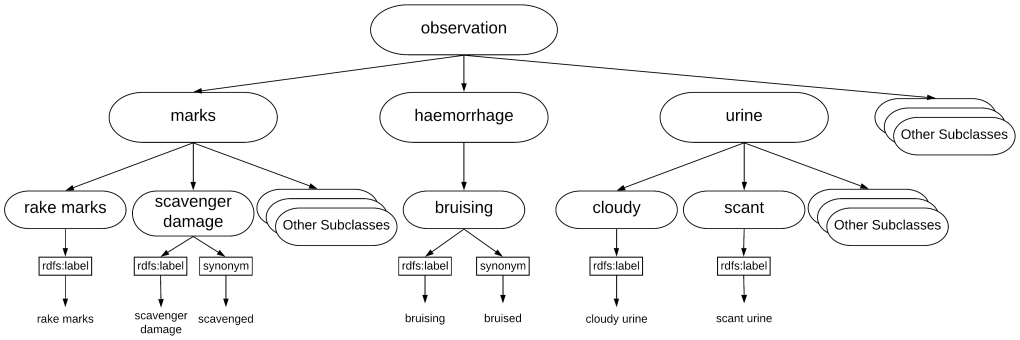


Figure 4: Structure of the observation ontology demonstrated using the marks, haemorrhage and urine class examples.

that precede causes of death. They were then categorised based on semantic similarity. For example, the "physical trauma" class represent cases where there is evidence of blunt force or penetration to the skin which appear to have been detrimental to the animal's health. This category captures conditions such as boat strikes, bottlenose dolphin attack trauma and entanglement (where rope, line or netting has wrapped around the animal).

3.3.3 Anatomy Ontology. Finally, the anatomy ontology was created based on the anatomical terms which were used to convey observations within the reports. The highest level of the "anatomical region" tree contains classes which relate to different organ systems within the body, or anatomical regions which are semantically linked. The latter situation applies, for example, to the "integument region" (relating to the skin) and the "external region" which mostly refers to external observations out with the main scope of those captured in the integument class.

The next level of subclasses generally represents different types of these regions. The decision was taken to make a distinction between having a parent-child relationship and an "isPartOf" attribute: it is not accurate, for example, to represent regions such as "the left valve of the heart" as a subclass of "heart", but it is still desirable to capture the relationship between these two regions. The "isPartOf" object property is transitive and asymmetric. This allows for instances such as the duodenum to be more accurately represented. The duodenum "isPartOf" the small intestine, and the small intestine "isPartOf" the intestines. Therefore, we can deduce that the duodenum "isPartOf" the intestines also. The structure of this ontology is shown in Figure 6, which shows some example anatomies in the ailmentary system. The anatomy ontology was manually populated and structured based on the headings used for sections of gross pathology reports (shown in Figure 2).

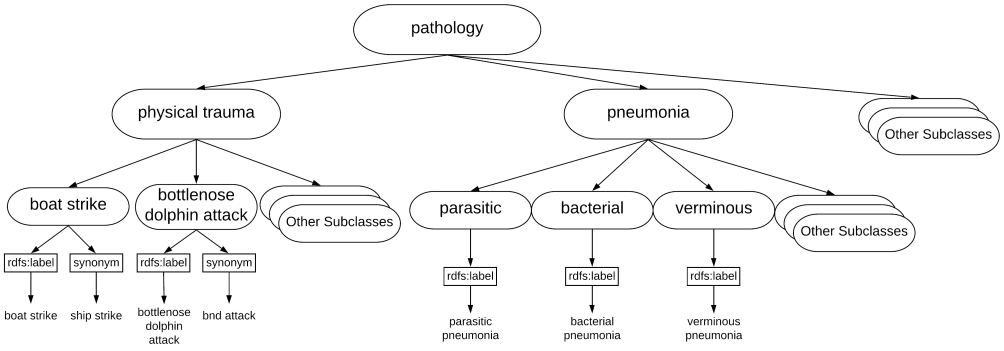


Figure 5: Structure of the pathology ontology demonstrated using the physical trauma and pneumonia class examples.

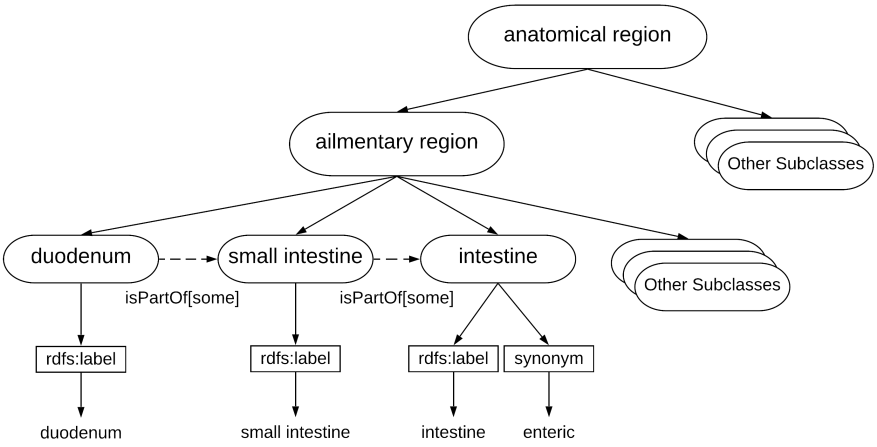


Figure 6: Hierarchical structure pf anatomy ontology demonstrated using example ailmentary region subclasses.

3.4 Information Retrieval

Information retrieval consists of several individual components within the framework including, *(a)* lexicon, *(b)* entity-relationship extraction engine, *(c)* anatomy, observation and presence recognition and *(d)* formatting to extract anatomical features, observations and pathologies. This pipeline is shown in the flowchart in Figure 1 as well as Algorithm 1.

3.4.1 Lexicons. Our anatomy, pathology and observation ontologies (refer Section 3.3) are used to identify entities. Two lexicons of key terms are generated by parsing the three ontology xml files and extracting their "rdfs:label", and "synonym" attributes. The observation lexicon was created using the observation and pathology ontologies, while the anatomy lexicon was created using the anatomy ontology. Pathological terms are incorporated in the observation lexicon because they can used to both represent the pathology of the specimen as a whole, and the condition of anatomical region. An example would be a case of physical trauma cause by entanglement which is a subclass of physical trauma within the pathology ontology. The inclusion of the term "entangled" could be treated as both an observation and pathology based on the representation within our ontology.

3.4.2 Entity-Relationship Extraction Engine. Reports are first segmented at the sentence level. As shown in Figure 2, sections are not always delimited by a full stop as one would expect. As such, we assume that sentences could also be delimited by new line characters ('\n'). Word level tokenisation is also performed. Words and punctuation are tagged using NLTK's POS Tagger library [14]. Words are then grouped together through "Noun Phrase Chunking" (NP-chunking). Empirically, we developed a simple feature-based grammar of tag patterns which represent entities and entity-relationships. Our grammar is passed into NLTK's *Regexparser* library to create chunks of entities and entity-relationships. The regular expression based grammar we defined for this task can be seen below.

```
NP: {<DT>?<JJ>*<VB.*>*<JJ>*<NN.*>+}
NP: {<NP><CC><NP>}
NP: {<VBD|VBN>}
NP: {<CD><RB>}
```

```
NP: {<NP><NP>}
IN: {<IN>}
REL:{<NP><IN><NP>}
REL:{<NP><TO><NP>}
REL:{<NP><:><NP>}
```

Entity chunks are grouped together as noun phrases (NP). The first rule captures any case where there is at least one noun preceded by any adjectives (JJ) or verbs (VB) and may include a determiner ("the", "a" etc.) denoted by 'DT'. If any past tense or past participle verbs are used separately, they is also chunked as a noun phrase to account for cases such as "right eye: scavenged". Lastly, NPs can be linked into a single NP where they are separated by coordinating conjunction terms such as 'and'.

NPs are then linked together into relationship (REL) chunks based on several conditions. Simple adjacency of two NPs is the first relationship as proximal entities are likely to relate. Prepositions (*e.g.*, 'in') were also of particular interest as they represent a relationship between that which precedes and follows them. The word 'to' is another good link between NPs given that phrases such as "damage to left flank" are very common. Lastly, we use the colon to capture cases where the anatomical entity is stated, then observations follow. An example of this is shown in Figure 2: "Blubber: NAD". This grammar is designed to capture relatively simple expressions, but can be expanded to incorporate more complex entity-relationships in future.

3.4.3 Anatomy, Observation and Presence Recognition. For each sentence in a report, each REL chunk is parsed and compared to the anatomical and observational lexicons. Where a NP chunk contains a sub-string that occurs in either lexicon, it is identified as anatomy or observation accordingly.

The implementation deliberately only incorporates NP - NP relationships (as defined in our grammar) as when only the relation subsection of the sentence is used for the marking of negated terms, one reduces the number of falsely negated terms. This means a relatively simple process for identifying negated words can be used, to a high degree of accuracy.

We use the NLTK mark_negated package for this purpose. The package adds '_NEG' as a suffix to any word between a negation and certain punctuation marks. For each REL chunk with an identified observation, a negated version of the statement is generated. The NP chunk containing the free text representation of the observation is compared to the same chunk after negated terms are marked. In the event an observational term is negated, it is considered to be an absent case. An example would be "no obvious rake marks on flank", where rake marks would be identified as an observational entity. When compared to the negation marked version of the text ("there are no obvious_NEG rake_NEG marks_NEG on_NEG flank_NEG") the negation of the observation would become apparent. In this event "rake marks" would be identified as "absent". The benefit of this approach is that one only marks negated terms at the relationship level. If one were to mark at the sentence level, unrelated negated terms would incorrectly cause for a classification of absence rather than presence.

When a recognised observational term is not attributed to an anatomical entity, it is still recorded as either present or absent and is not attributed to an anatomical region. There are a number of reasons why an anatomical entity might not be identified: the term used is not represented within the anatomical ontology; the observation is not used in relation to an anatomical entity; or the grammar used for chunking fails to capture relevant NP chunks within a relationship.

3.4.4 Formatting Findings. The information extracted is summarised in a dictionary implemented in Python, which can then be used for analysis or classification systems. The anatomy and observation terms are represented as strings, while the presence or absence of an observation is stored as a Boolean value. Some examples are shown below:

```
{
    'anatomy': 'right pectoral fin',
    'observation': 'scavenger damage',
    'presence': True
}
{
    'anatomy': 'epidermis',
    'observation': 'rake marks',
    'presence': False
}
{
    'anatomy': 'skull',
    'observation': 'nad',
    'presence': True
}
```

4 USE CASE, RESULTS AND ANALYSIS

To analyse the effectiveness of our approach, we identified observations which could negate or confirm a chosen pathological finding, and used the presence or absence of these observations as a means of classification. Bottlenose Dolphin attacks (BDA) on harbour porpoises are a very common cause of death within the dataset, with 50 of the 193 cases listing BDA as the key finding in the SMASS database.

A deterministic classifier was created using extracted empirical observations. The classes chosen were either explicit mentions of BDA, or strong indicators such as "rake marks". Where BDA is mentioned, it's absence or presence is sufficient to classify the case as "Non-BDA" or "BDA". The observation "rake marks", however, can also be used to describe some grey seal attacks (GSA). As such, we then filter all observations relevant to seal attacks and claw marks (an indicator of a GSA). We make the assumption that if there is evidence of both a GSA and a BDA, an explicit mention of BDA should be found. If the document has not been classified using these rules, presence of "rake marks" results in "BDA" classification. The deterministic classifier's sequence of decisions is shown in Algorithm 2. The cause of death stored in the SMASS database was used as ground truth for classifier evaluation.

The results for the Bottlenose Dolphin attacks use case are shown in Table 1. The approach achieved an overall accuracy of 83.4% and an F1-score of 0.83. BDA classification achieved a precision of 0.70, a recall of 0.64, and F1-score of 0.67. Non-BDA classification achieved a precision of 0.88, a recall of 0.90, and F1-score of 0.89. Of the 193 reports used, 50 were cases of BDA and 143 were Non-BDA cases based on the cause of death stored in the SMASS database.

Algorithm 2: Deterministic BDA classification process based on presence or absence of observations.

```
Result: prediction
if Any observation is a BDA term then
    if observation present then
        prediction <- "BDA";
    else
        prediction <- "Non-BDA";
    end
    return prediction;
else
    if Any present observation is a GSA or claw mark term
     then
        prediction <- "Non-BDA";
        return prediction;
    else
        if Any present observation is a rake mark then
            prediction <- "BDA";
            return prediction;
        end
        prediction <- "Non-BDA";
        return prediction;
    end
end
```

Metrics	Cumulative	BDA	Non-BDA
Accuracy	0.83	-	-
ROC-AUC	0.77	-	-
Recall	-	0.64	0.90
Precision	-	0.70	0.88
F1-score	0.83	0.67	0.89
Support	193	50	143

Table 1: BDA classifier performance evaluation metric scores.

A Receiver Operating Characteristic (ROC) curve (Figure 7) was generated using the BDA precision and recall values listed above which achieved the Area Under Curve (AUC) score of 0.77. The disadvantage of labelling based on cause of death is that there are many instances where a BDA has occurred but a separate finding has been identified as the cause of death. This leads to some false positive (FP) classifications as BDA terms and indicators are still described. This can be seen in the confusion matrix in Figure 8. Given the deterministic nature of our classifier, there are three possible causes of incorrect classifications. The first is the presence of a separate more significant finding which caused death, even though a BDA occurred; the second is the use of an significant term outwith the scope it was intended; and the third is that a significant finding is not successfully identified by the entity-relationship engine.

When analysing cases of FPs, some statements such as "rake marks, assumed bird" lead to an incorrect detection. This reflects

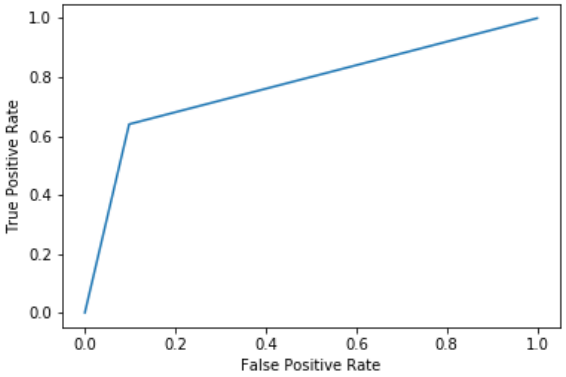


Figure 7: ROC-AUC curve of BDA classifier predictions. AUC = 0.771

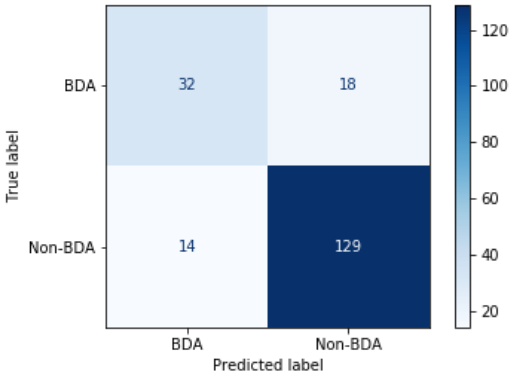


Figure 8: Confusion matrix of BDA classifier predictions.

the difficulty of varying terminology usage between pathologists. Several FPs included instances where BDA rake marks were "healed" or "healing". Some false negatives (FNs) were caused by the lack of explicit mentions of BDAs and there being "no obvious rake marks". This suggests that there were other indicators of BDA despite the absence of rake marks.

The recall (0.64) and precision (0.70) scores of the BDA classifications are relatively low due to the simplicity of the feature-based grammar, and the total number of true positives (TPs) being underestimated. This being said, analysis of FPs and FNs also showed some cases where significant phrases were not captured by the feature-based grammar in the entity-relationship engine.

The precision and recall scores associated with Non-BDAs (0.88 and 0.90 respectively) are considerably higher. When analysed, it was found that several cases were correctly identified as Non-BDA due to the exclusion of GSAs in the deterministic classifier. This shows that the creation of an inclusion/exclusion based determiner can be used to increase trust in positive classifications, meaning any insight obtained is more robust to scrutiny. While the classifier had

minor shortcomings due to the grammar used, the results are very promising for future work. Using cause of death as a label leads to lower performance metrics than anticipated. The use of a manually labelled dataset would naturally produce more realistic results, but in its absence, we can still get a good understanding of the classifier characteristics.

Another thought to consider is that one necropsy report contains many other fields pertaining to morphology, confidence in diagnosis and other free text sections such as the histopathology report and conclusion sections. By incorporating relevant fields and applying a similar information retrieval process to other free text sections, more accurate, complex, and inclusive determiners would be defined, meaning a higher confidence in positive or negative results.

5 CONCLUSIONS

We proposed Ir-Man, an information retrieval framework for marine animal necropsy analysis. The framework applied and adapted information retrieval techniques to reports in a previously unexplored domain. Necropsy reports of stranded marine mammals provide a unique insight into marine ecology; the ability to access and aggregate this information will allow for more useful epidemiological analysis. Despite the challenges associated with mining semi-structured gross pathology reports, our information retrieval framework achieved a baseline accuracy of 83.4% when classifying BDAs on harbour porpoises. Future work will include the incorporation of more complex feature-based grammar representations which would identify structure within text more effectively; expanding the ontologies to incorporate other cetaceans; and defining and detecting further nuance between different observation classes. Most importantly, the framework will be used to further pathological and epidemiological understanding within the marine mammal domain.

ACKNOWLEDGEMENT

We acknowledge the support of a joint PhD studentship by University of Stirling and Scotland's Rural College (SRUC).

REFERENCES

[1] Olivier Bodenreider. 2004. The Unified Medical Language System (UMLS): integrating biomedical terminology. *Nucleic acids research* 32 Database issue (2004), D267–70.
[2] Nathan Bollig, Lorelei Clarke, Elizabeth Elsmo, and Mark Craven. 2020. Machine learning for syndromic surveillance using veterinary necropsy reports. *PloS one* 15, 2 (2020), e0228105.
[3] Wendy W Chapman, Will Bridewell, Paul Hanbury, Gregory F Cooper, and Bruce G Buchanan. 2001. A simple algorithm for identifying negated findings and diseases in discharge summaries. *Journal of biomedical informatics* 34, 5 (2001), 301–310.
[4] Albert Comelli, Luca Agnello, and Salvatore Vitabile. 2015. An ontology-based retrieval system for mammographic reports. In *IEEE Symposium on Computers and Communication (ISCC)*. IEEE, 1001–1006.
[5] Sean Falconer. 2010. Ontograf protege plugin. *Place: Available at: http://protegewiki. stanford. edu/wiki/OntoGraf [Accessed: 21/03/2014]* (2010).
[6] Jeffrey Friedlin, Malika Mahoui, Josette Jones, and Patrick Jamieson. 2011. Knowledge discovery and data mining of free text radiology reports. In *IEEE First International Conference on Healthcare Informatics, Imaging and Systems Biology*. IEEE, 89–96.
[7] Lenz Furrer, Susanne Küker, John Berezowski, Horst Posthaus, Flavie Vial, Fabio Rinaldi, Thierry Poibeau, and Pamela Faber. 2015. Constructing a syndromic terminology resource for veterinary text mining. (2015).
[8] Hongyuan Gao, Erin J Aiello Bowles, David Carrell, and Diana SM Buist. 2015. Using natural language processing to extract mammographic findings. *Journal of biomedical informatics* 54 (2015), 77–84.
[9] Zelalem Gero and Joyce C Ho. 2019. NamedKeys: Unsupervised Keyphrase Extraction for Biomedical Documents. In *Proceedings of the 10th ACM International Conference on Bioinformatics, Computational Biology and Health Informatics*. 328–337.
[10] Lejun Gong, Jie Yan, Jiacheng Feng, and Ronggen Yang. 2015. A dictionary-based approach to identify biomedical concepts. In *12th International Conference on Fuzzy Systems and Knowledge Discovery (FSKD)*. IEEE, 1091–1095.
[11] Lejun Gong, Ronggen Yan, Quan Liu, Haoyu Yang, Gene Yang, and Kaiyu Jiang. 2016. Extraction of biomedical informtion related to breast cancer using text mining. In *12th International Conference on Natural Computation, Fuzzy Systems and Knowledge Discovery (ICNC-FSKD)*. IEEE, 801–805.
[12] Donghyeon Kim, Jinhyuk Lee, Chan Ho So, Hwisang Jeon, Minbyul Jeong, Yonghwa Choi, Wonjin Yoon, Mujeen Sung, and Jaewoo Kang. 2019. A neural named entity recognition and multi-type normalization tool for biomedical text mining. *IEEE Access* 7 (2019), 73729–73740.
[13] Susanne Küker, Celine Faverjon, Lenz Furrer, John Berezowski, Horst Posthaus, Fabio Rinaldi, and Flavie Vial. 2018. The value of necropsy reports for animal health surveillance. *BMC veterinary research* 14, 1 (2018), 191.
[14] Edward Loper and Steven Bird. 2002. NLTK: the natural language toolkit. *arXiv preprint cs/0205028* (2002).
[15] Vajenti Mala and DK Lobiyal. 2015. Concepts extraction for medical documents using ontology. In *International Conference on Advances in Computer Engineering and Applications*. IEEE, 773–777.
[16] Sebastian Mate, Felix Köpcke, Dennis Toddenroth, Marcus Martin, Hans-Ulrich Prokosch, Thomas Bürkle, and Thomas Ganslandt. 2015. Ontology-based data integration between clinical and research systems. *PloS one* 10, 1 (2015).
[17] Christopher J Mungall, Carlo Torniai, Georgios V Gkoutos, Suzanna E Lewis, and Melissa A Haendel. 2012. Uberon, an integrative multi-species anatomy ontology. *Genome biology* 13, 1 (2012), R5.
[18] Shamkant B Navathe. 2007. Text Mining and Ontology Applications in Bioinformatics and GIS. In *Sixth International Conference on Machine Learning and Applications (ICMLA 2007)*. IEEE, xviii–xix.
[19] Sarah E Nelms, James Barnett, Andrew Brownlow, NJ Davison, Rob Deaville, Tamara S Galloway, Penelope K Lindeque, D Santillo, and Brendan J Godley. 2019. Microplastics in marine mammals stranded around the British coast: ubiquitous but transitory? *Scientific Reports* 9, 1 (2019), 1–8.
[20] Natalya Fridman Noy, Monica Crubézy, Ray W Fergerson, Holger Knublauch, Samson W Tu, Jennifer Vendetti, and Mark A Musen. 2003. Protégé-2000: an open-source ontology-development and knowledge-acquisition environment.. In *AMIA... Annual Symposium proceedings. AMIA Symposium*, Vol. 2003. American Medical Informatics Association, 953–953.
[21] P Sudeshna, S Bhanumathi, and MR Anish Hamlin. 2017. Identifying symptoms and treatment for heart disease from biomedical literature using text data mining. In *International Conference on Computation of Power, Energy Information and Commuincation (ICCPEIC)*. IEEE, 170–174.
[22] Hao Wei, Mingyuan Gao, Ai Zhou, Fei Chen, Wen Qu, Chunli Wang, and Mingyu Lu. 2019. Named entity recognition from biomedical texts using a fusion attention-based BiLSTM-CRF. *IEEE Access* 7 (2019), 73627–73636.
[23] Rosie Williams, Mariel ten Doeschate, David J Curnick, Andrew Brownlow, Jonathan L Barber, Nicholas J Davison, Robert Deaville, Matthew Perkins, Paul D Jepson, and Susan Jobling. 2020. Levels of Polychlorinated Biphenyls Are Still Associated with Toxic Effects in Harbor Porpoises (Phocoena phocoena) Despite Having Fallen below Proposed Toxicity Thresholds. *Environmental Science & Technology* 54, 4 (2020), 2277–2286.
[24] Rosie S Williams, David J Curnick, Jonathan L Barber, Andrew Brownlow, Nicholas J Davison, Rob Deaville, Matthew Perkins, Susan Jobling, and Paul D Jepson. 2020. Juvenile harbor porpoises in the UK are exposed to a more neurotoxic mixture of polychlorinated biphenyls than adults. *Science of the Total Environment* 708 (2020), 134835.
[25] Yujie Yang, Yunpeng Cai, Wenshu Luo, Zhifeng Li, Zhenghui Ma, Xiaolu Yu, and Haibo Yu. 2013. An ontology-based approach for text mining of stroke electronic medical records. In *IEEE International Conference on Bioinformatics and Biomedicine*. IEEE, 288–291.
[26] Sendong Zhao, Meng Jiang, Ming Liu, Bing Qin, and Ting Liu. 2018. CausalTriad: toward pseudo causal relation discovery and hypotheses generation from medical text data. In *Proceedings of the 2018 ACM International Conference on Bioinformatics, Computational Biology, and Health Informatics*. 184–193.

Appendix 4 - Analysing hyperplasia in Atlantic salmon gills using empirical wavelets

Analysing hyperplasia in Atlantic salmon gills using empirical wavelets

Alexander F. B. Carmichael[a,b], Johanna L. Baily[a], Aaron Reeves[c], Gabriela Ochoa[a], Annette S. Boerlage[b], George Gunn[b], Rosa Allshire[a], and Deepayan Bhowmik[d]

[a]University of Stirling, Stirling, UK
[b]Scotland's Rural College (SRUC), Inverness, UK
[c]Center for Applied Public Health Research, RTI International, USA
[d]Newcastle University, Newcastle upon Tyne, UK

ABSTRACT

Measuring hyperplasia in Atlantic salmon gills can give important insight into fish health and environmental conditions such as water quality. This paper proposes a novel histology image classification technique to identify hyperplastic regions using an emerging signal decomposition technique, Empirical Wavelet Transform (EWT) in combination with a fully connected neural network (FCNN). Due to its adaptive nature, we hypothesise and show that EWT effectively represents unique features of gill histopathology whole slide images that help in the classification task. Our hybrid approach is unique and significantly outperformed regular deep learning-based methods considering a joint speed-accuracy metric.

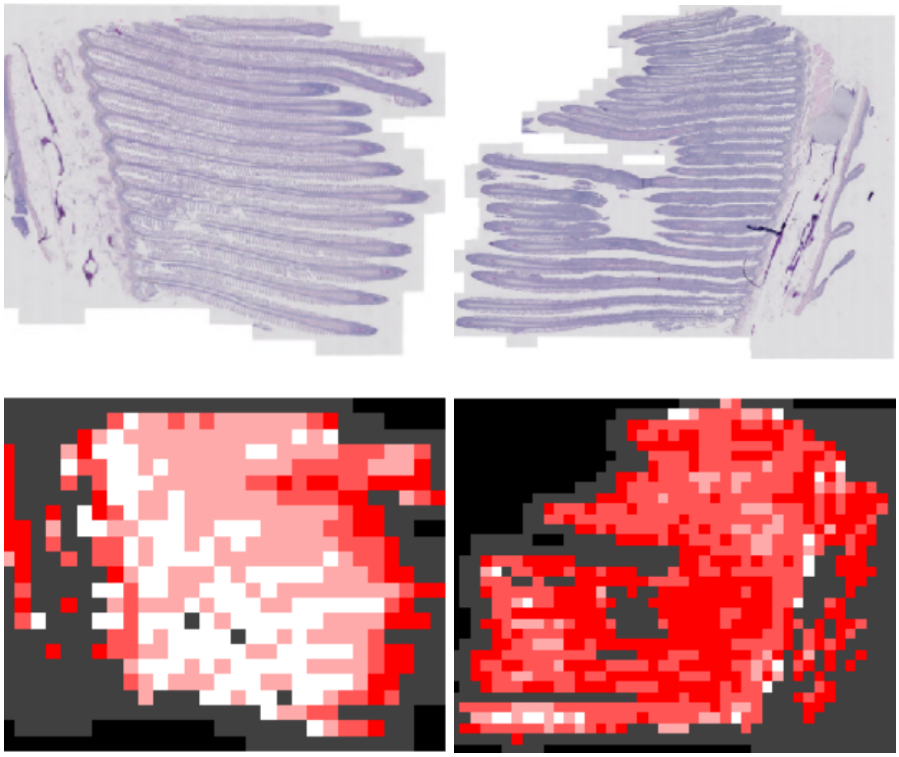
Keywords: Hyperplasia, empirical wavelet transform, image classification, histopathology, deep learning, Atlantic salmon gills, digital pathology

1. INTRODUCTION

Tissue samples taken from fish during post-mortem analysis can give valuable insight into the animal's condition at the point of death. An essential aspect of this is histopathology, which involves assessing changes in tissue at the microscopic level. Epithelial hyperplasia is a lesion of cell proliferation and a strong indicator of local irritation, infectious disease, or poor water quality. This paper proposes a new classification approach that assesses the severity of hyperplasia in Atlantic Salmon whole-slide images (WSIs). We have incorporated pathological domain expertise in the image processing pipeline and propose a novel feature engineering technique based on Empirical Wavelet Transform (EWT), it's subband statistics and a fully connected neural network (FCNN). To identify local affected regions, a tile-based method is used that allows the pathologists to visualise and understand which regions of a WSI contribute to an overall classification. An example output of the proposed algorithm is shown in Fig. 1.

Hyperplasia in gills has a significant impact on the morphology of the microscopic anatomy. Epithelial hyperplasia can be caused by local irritation, infectious disease, or poor water quality. Tissue samples taken from healthy gills form a comb-like structure where the comb's teeth, known as primary lamellae, further branch into secondary lamellae (Figures 1 & 2). In a healthy gill, the secondary lamellae consist of a linear red blood cell-filled channel surrounded by a thin layer of cells known as the epithelium. In mild and moderate cases of

Further author information: (Send correspondence to A.F.B.C.)
A.F.B.C.: E-mail: a.f.carmichael@stir.ac.uk
J.L.B.: E-mail: j.l.baily@stir.ac.uk
A.R.: E-mail: areeves@rti.org
G.O.: E-mail: gabriela.ochoa@stir.ac.uk
A.S.B.: E-mail annette.boerlage@sruc.ac.uk
G.G.: E-mail: george.gunn@sruc.ac.uk
R.A.: rosa.allshire@stir.ac.uk
D.B.: E-mail deepayan.bhowmik@newcastle.ac.uk



(a) Normal/Mild (b) Severe

Figure 1. A heatmap of our approach applied to gill WSIs. White indicates healthy and red indicates severe hyperplasia.

hyperplasia, the number of cells in the epithelial layer increase, and as such, the space between secondary lamellae decreases. In a severe case, the secondary lamellae fuse entirely (Fig. 2). High degrees of hyperplasia diminish the gill's respiratory function and ability to extract oxygen from water, meaning that measuring hyperplasia in gills is important for understanding the overall health of the fish.

Classification of hyperplasia in WSIs can be automated through frequency domain texture analysis. However, traditional image decomposition methods either rely on fixed template-based (e.g., wavelets) or data-driven (e.g., Empirical Mode Decomposition (EMD)[1]) approaches. While the former lacks flexibility due to rigid template structures, the latter lacks an underlying mathematical basis. A recent development[2,3] attempted to address the gap by proposing empirical wavelet transform which is adaptive to the input signal. We hypothesise that content adaptive signal decomposition along with an artificial neural network is advantageous for texture understanding and classification in analysing gill WSIs.

In this paper, we propose a tile-based classification framework allowing for statistic generation. An approach that quantifies the severity of a variety of lesions in an individual WSI can create a profile of an animal that could complement other pathological or epidemiological data. As such, our work makes the following contributions:

- A unique data-driven approach for hyperplasia analysis in gill histology images that does not currently exist, which can also be used to produce visualisations.
- A new parametric feature generation method (in combination with an FCNN) that translates expert knowledge into mathematically explainable features using both non-adaptive wavelets and EWT.
- A scale and rotation agnostic method of representing image features allowing for effective models to be developed using small datasets.

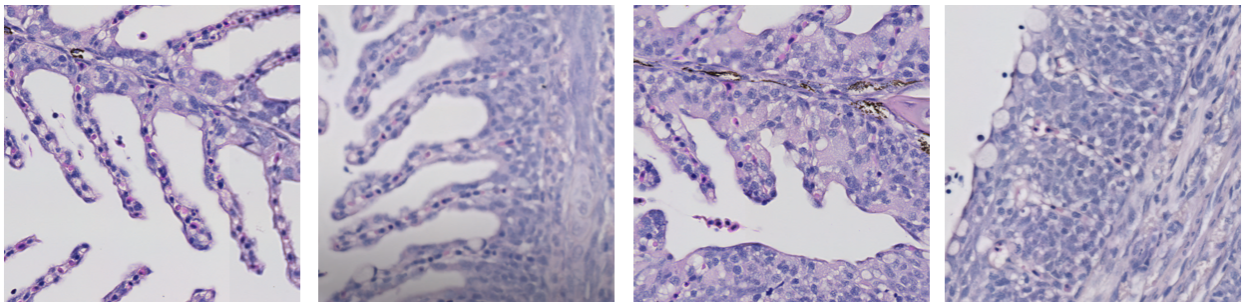


Figure 2. Tiles extracted from gill WSIs. From left to right - Normal, Mild, Moderate, and Severe cases of hyperplasia.

2. RELATED WORK

Machine learning (ML) is being increasingly used in aquaculture to optimise fish growth and performance.[4] Similarly, ML is revolutionising histopathology understanding of many tissue types and animal species, most notably in human medicine[5] and oncology.[6] Humans have naturally been the focus of much of this research, given its importance in tasks like tumour detection.[7] CNNs, in particular, have seen wide usage for applications such as brain tumour segmentation,[8] gastrointestinal cancer classification,[9] and lung cancer classification.[10]

Very little work exists for the automated histological analysis of gill tissue in fish despite heavy reliance on such tools in the field. Current methods use traditional image pre-processing steps and alternative staining techniques for quantifying mucous cells in salmon skin[11] or to study gill health.[12] Jayasuriya developed and used a tool that evaluated morphological changes in salmon gills to produce descriptors in individual cases automatically.[13] While useful for in-depth analysis of individual gill WSIs, a data-driven approach for analysis across large numbers of WSIs is desirable. Template-based computer vision pipelines are more challenging to develop than data-driven machine learning-based approaches. Sveen et al. have successfully used deep learning to automatically segment Atlantic Salmon skin tissue;[14] however, there has been little to no progress in automated analysis of gill histology images.

The EWT has been applied to glaucoma classification in Fundus Imaging. However, glaucoma classification presents a much more consistent morphological structure than gill WSIs. Primary and secondary lamellae in gills vary in structure based on location, and branches may overlap. Furthermore, the tissue sectioning process frequently introduces irrelevant artefacts. Maheshwari et al. used 2D empirical Littlewood-Paley (LP) wavelet subband images to obtain correntropy features to identify cases of glaucoma in fundus images.[15] Similarly, Kirar et al. developed an EWT-based approach that combined traditional discrete wavelet transforms with the 2D LP-EWT to produce features from subband images, including Zernike moment, Hu's Invariant Moments, chip histogram, and grey level matrix approaches.[16] Chaudhary et al.[17] used a Fourier-Bessel series expansion-based EWT (2D-FBSE-EWT) and transfer learning with four pre-trained Resnet-50 models to evaluate subband images.

3. METHODOLOGY

Our approach consists of three main sections: preprocessing, the subband image generation using 2D LP-EWT, and parametric feature calculation. These steps are shown in Fig. 3. The vectors generated are then used as features for machine learning models. Our approach aims to incorporate expert knowledge as model features. In this context, ground truth has been created in consultation with the expert pathologist identifying several components of gill histology slides which help differentiate between healthy/normal, mild, moderate, and severe cases of hyperplasia. Indicators of hyperplasia were summarised as: an increase in overall tissue area; a decrease in space between secondary lamellae, and a shift in tissue colour.

We evaluated potential features and included or excluded them based on how much they represent the characteristics listed above. By applying the EWT to separate RGB channels of the tiles, we represent the colour and overall texture of the image. Before we applied the EWT to each tile, a series of preprocessing

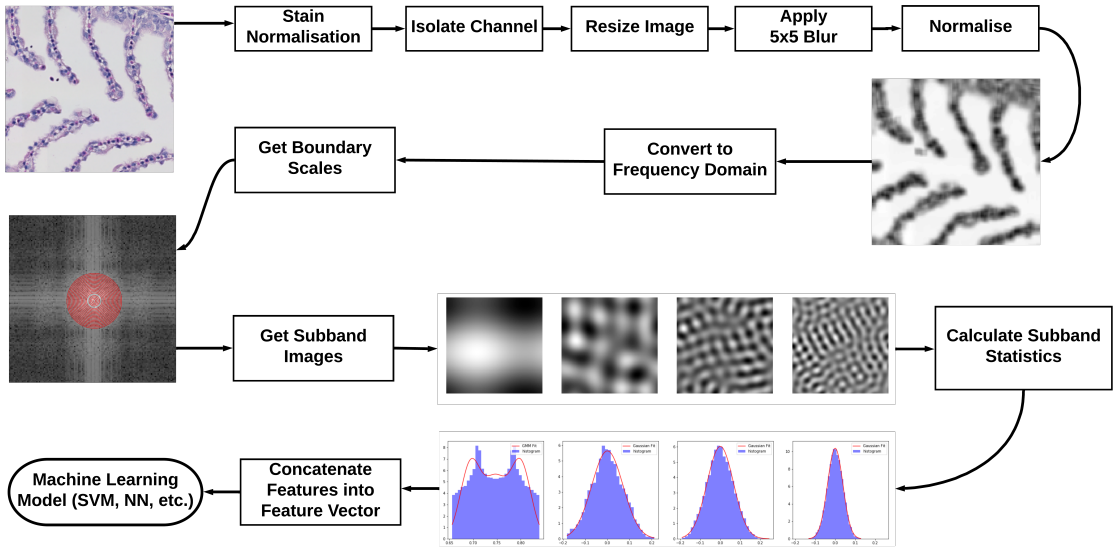


Figure 3. A flowchart showing the main sections of our approach: preprocessing, EWT, and parametric feature generation. A Gaussian Mixture Model fitted to the lowest subband is shown, as well as Gaussian distributions for three other subbands.

operations were applied. First, we resized each image from 1024×1024 to 128×128 using pixel area relation resampling. We then used a low-pass 5×5 average blurring filter before normalisation.

3.1 Littlewood-Paley Empirical Wavelet Transform

The traditional 2D LP transform involves filtering images using 2D wavelets in the Fourier domain, with annuli supports around the origin. Scales are divided by these supports, the radii of which are calculated using the dyadic decomposition of the Fourier plane. Gilles et al. proposed a version of this method that involved the empirical detection of annuli radii.[3] The 2D LP-EWT, $\mathcal{W}^{\mathcal{ELP}}$, of an input image, f, is described in Eq. (1), which is denoted by the 2D Fourier boundary detection function, $\mathcal{F}_2$, its inverse, $\mathcal{F}_2^*$, and the wave function, ψ. $\mathbf{x} = (x_1, x_2)$ is the position in the 2D plane while $\omega = (\omega_1, \omega_2)$ is the position in the frequency domain. In our work we empirically chose n=20, meaning 20 subbands are produced. The central panel in Fig. 3 shows the frequency domain image, the boundaries generated, and the resultant subband images.

$$\mathcal{W}_f^{\mathcal{ELP}}(n, \mathbf{x}) = \mathcal{F}_2^* \left(\mathcal{F}_2(f)(\omega) \overline{\mathcal{F}_2\left(\psi_n\right)(\omega)} \right). \tag{1}$$

Naturally, the boundary detection method used to build the filter bank, $\mathcal{B}$, described previously, is fundamental to the LP-EWT approach. The first step is to perform a Pseudo-Polar Fast Fourier Transform (FFT), $\widetilde{\mathcal{F}}_P(|\omega|)$, of the original image based on the number of defined filters. The average is then calculated with regard to each discrete angle, θ, in the 1D Fourier spectrum, the number of which is defined as N_θ. This process is described in Eq. (2).

$$\widetilde{\mathcal{F}}_P(|\omega|) = \frac{1}{N_\theta} \sum_{i=0}^{N_\theta - 1} \left| \mathcal{F}_P(f)\left(\theta_i, |\omega|\right) \right|. \tag{2}$$

The filter bank is then constructed using the spectral radii obtained, shown in Eq. (3). More details can be found in Gilles et al.'s 2014 paper on 2D empirical wavelets.[3] The outcome of this process is the boundaries shown in Fig. 3.

$$\mathcal{B} = \left\{ \phi_1(\mathbf{x}), \{\psi_n(\mathbf{x})\}_{n=1}^{N-1} \right\}. \tag{3}$$

3.2 Parametric Feature Engineering

Given that we identified colour and texture as relevant characteristics for our approach to capture, the LP-EWT method described above was applied to separate colour channels of each tile. We applied a parametric method to the subband images generated by EWT to produce the features used for model training and testing.

Our approach produces, n, subband images when we apply LP-EWT to the tile. We then extract statistical features from each image by fitting bespoke probability distribution functions (PDF) to flattened representations of the tiles' pixels. A Gaussian Mixture Model (GMM) PDF[18] is fitted to the 1st image, with a given number of, $\mathcal{K}$, components. For each k^{th} component, a mean, μ_k, a variance, σ_k, and a weight, ϕ_k are produced. In our approach, we have empirically set $\mathcal{K}$ to be 3, meaning for the 1st subband image we obtain 9 features. Equation 4 describes the GMM PDF-based feature generation process.

$$p(x) = \sum_{i=1}^{K} \phi_i \mathcal{N}\left(x \mid \mu_i, \sigma_i\right), \tag{4}$$

where,

$$\mathcal{N}\left(x \mid \mu_i, \sigma_i\right) = \frac{1}{\sigma_i\sqrt{2\pi}} \exp\left(-\frac{(x-\mu_i)^2}{2\sigma_i^2}\right),$$

such that,

$$\sum_{i=1}^{K} \phi_i = 1.$$

For all subband images except the first, a Gaussian PDF is fitted based on the pixel values (shown in Eq. 5). The σ is kept and used for the feature vector, while the mean, μ, is discarded as $\mu = 0$ in all cases.

$$p(x; \mu, \sigma) = \frac{1}{\sigma\sqrt{2\pi}} e^{-\frac{1}{2}\left(\frac{x-\mu}{\sigma}\right)^2}. \tag{5}$$

This process is applied to three different representations of the tile using the red, green, and blue colour channels. Given that n=20 in this work, we generate 29 features for each colour channel and 87 in total.

The features extracted from the 1st subband image using the GMM PDF are concatenated and organised.

The features extracted from the other subband images from each channel $(\mathcal{H}_r, \mathcal{H}_g, \mathcal{H}_b)$ are then also concatenated to produce the feature vector, $\mathcal{H}$:

$$\mathcal{H} = (\mathcal{H}_r, \mathcal{H}_g, \mathcal{H}_b), \tag{6}$$

where,

$$\begin{aligned}
\mathcal{H}_r &= (\mathcal{M}_r, \sigma_{r1}, \sigma_{r2}, ...\sigma_{rn}),\\
\mathcal{H}_g &= (\mathcal{M}_g, \sigma_{g1}, \sigma_{g2}, ...\sigma_{gn}),\\
\mathcal{H}_b &= (\mathcal{M}_b, \sigma_{b1}, \sigma_{b2}, ...\sigma_{bn}).
\end{aligned}$$

$$\mathcal{M} = (\mu_0, ..., \mu_k, \sigma_0, ..., \sigma_k, \phi_0, ..., \phi_k). \tag{7}$$

The features extracted from the other subband images from each channel $(\mathcal{H}_r, \mathcal{H}_g, \mathcal{H}_b)$ are then also concatenated to produce the feature vector, $\mathcal{H}$:

The feature vector $(\mathcal{H})$ is then passed to an FCNN with 3 hidden layers of 1024 neurons.

Algorithm 1: Parametric feature generation

T: Tile image. P: Predicted class.
$n \leftarrow 20$; // n = number of subband images generated by EWT.
$features.append(tissue\_area(T))$; // Get tissue area feature.
$channels \leftarrow separate\_rgb(T)$; // Get individual channels.
for $channel \in channels$ **do**
  $channel \leftarrow pre\_processing(channel)$; // Resize to 128x128, apply 5x5 blur, and normalise.
  $subband\_images \leftarrow EWT(n, channel)$; // Get channel subband images using EWT boundaries.
  for $subband \in subband\_images$ **do**
    $pixels \leftarrow flatten(subband)$; // Get pixel values.
    $subband\_statistics \leftarrow distribution.fit(pixels)$; // Generate statistics from probability density function.
    $channel\_Features.append(subband\_statistics)$
  end
  $features.append(channel\_features)$; // Add channel statistics to feature vector.
  $P \leftarrow model.predict(features)$; // Model predicts Tile class.
end

4. EXPERIMENTAL RESULTS AND DISCUSSION

4.1 Dataset Preparation

The dataset used in this work is composed of 5 WSIs of gills. We divided each WSI into tiles of size 1024x1024. Each tile was manually categorised as either *normal*, *mild*, *moderate*, *severe*, *other*, or *empty*. Any tile with an identifiable quantity of primary or secondary lamellae was labelled based on the severity of hyperplasia represented in the tile, with healthy lamellae classified as *normal* and tiles with hyperplasia classified as *mild*, *moderate*, or *severe* as appropriate. Tiles categorised as *other* (for-lamellae tissue) or *empty* were not used in the experiment. The dataset was edited and verified by an expert pathologist. The final version of the dataset consists of 1465 tiles. The moderate class is slightly underrepresented with only 169 samples. We applied Vahadane stain normalisation[19] to each tile in the dataset, with a representative large section of a withheld WSI used as a reference. This was to minimise the possibility that models may learn based on the staining characteristics of individual WSIs.

Model	Average Accuracy Across 10 Folds (%)	Fold Training and Testing Runtime w/ Preprocessing (Minutes:Seconds)
AlexNet[20]	58.84	17:12
ResNet18[21]	71.94	27:41
MobileNetV3-Small[22]	67.78	13:17
InceptionV3[23]	73.31	83:15
Biorthogonal 1.3	64.46	3.55
Haar	63.26	3:53
LP-EWT	**73.10**	**15:18**

Table 1. Average accuracy and runtime results across 10-folds. Our LP-EWT approach achieved a similar accuracy to the best-performing CNN evaluated while using less than a 5th of the total training and testing runtime.

4.2 Experimental Setup

We applied our parametric feature generation approach to both traditional wavelet methods (Biorthogonal 1.3 and Haar wavelet using PyWavelets[24]) and the LP-EWT. For the traditional wavelets, the 3 levels of subband images were used, with the GMM features extracted from the LL subband. We also evaluated several non-pretrained CNNs - AlexNet,[20] ResNet18,[21] MobileNetV3-Small,[22] and Inception V3.[23] We trained these models

using 10-fold cross-validation. For CNNs, we trained each fold over ten epochs with a batch size of 32. Due to the considerably shorter training times of the wavelet-based approaches, 50 epochs were used. The NNs used for the wavelet-based methods consisted of 3 layers of 1024 nodes, each with ReLU activation functions. They were trained using Cross Entropy Loss and an Adam optimiser.

4.3 Results and Discussion

We recorded accuracy scores, and training and testing runtimes for each model (Table 1). Of the models evaluated, InceptionV3 and LP-EWT achieved the highest accuracy scores of 73.31 and 73.10 respectively for the classification task. InceptionV3's total training and testing time across one fold was 83:15. In contrast, the LP-EWT had a one-fold training and testing time of 15:18. It should be noted that the results were generated using a small dataset, demonstrating the effectiveness of each approach on tasks where data availability is limited. We also created visualisations of WSIs using the LP-EWT to evaluate the effectiveness of our approach (shown in Figure 1). We did this to empirically compare the aggregate scores produced by our model and the ground truth provided by an expert pathologist. In both cases, our approach agreed with the expert; however, some non-lamellar tissue was included and classified as severe, despite not being relevant for the task of hyperplasia analysis. In the future, we intend to include a semantic segmentation step in the pipeline to identify regions relevant to the task of hyperplasia analysis.

ACKNOWLEDGMENTS

A.F.B.C's PhD studentship was joint funded by the University of Stirling, and Scotland's Rural College (SRUC). R.A's PhD was funded by Cooke Aquaculture and the University of Stirling.

REFERENCES

[1] Huang, N. E., Shen, Z., Long, S. R., Wu, M. C., Shih, H. H., Zheng, Q., Yen, N.-C., Tung, C. C., and Liu, H. H., "The empirical mode decomposition and the hilbert spectrum for nonlinear and non-stationary time series analysis," *Proceedings of the Royal Society of London. Series A: mathematical, physical and engineering sciences* **454**(1971), 903–995 (1998).

[2] Gilles, J., "Empirical wavelet transform," *IEEE Transactions on Signal Processing* **61**(16), 3999–4010 (2013).

[3] Gilles, J., Tran, G., and Osher, S., "2d empirical transforms. wavelets, ridgelets, and curvelets revisited," *SIAM Journal on Imaging Sciences* **7**(1), 157–186 (2014).

[4] Zhao, S., Zhang, S., Liu, J., Wang, H., Zhu, J., Li, D., and Zhao, R., "Application of machine learning in intelligent fish aquaculture: A review," *Aquaculture* , 736724 (2021).

[5] Nimbkar, S., Auddy, M., Manoj, I., and Shanmugasundaram, S., "Novel techniques for quality evaluation of fish: A review," *Food Reviews International* , 1–24 (2021).

[6] Zuraw, A. and Aeffner, F., "Whole-slide imaging, tissue image analysis, and artificial intelligence in veterinary pathology: An updated introduction and review," *Veterinary Pathology* , 03009858211040484 (2021).

[7] Carneiro, G., Zheng, Y., Xing, F., and Yang, L., "Review of deep learning methods in mammography, cardiovascular, and microscopy image analysis," in [*Deep Learning and Convolutional Neural Networks for Medical Image Computing*], 11–32, Springer (2017).

[8] Biratu, E. S., Schwenker, F., Ayano, Y. M., and Debelee, T. G., "A survey of brain tumor segmentation and classification algorithms," *Journal of Imaging* **7**(9), 179 (2021).

[9] Kuntz, S., Krieghoff-Henning, E., Kather, J. N., Jutzi, T., Höhn, J., Kiehl, L., Hekler, A., Alwers, E., von Kalle, C., Fröhling, S., et al., "Gastrointestinal cancer classification and prognostication from histology using deep learning: Systematic review," *European Journal of Cancer* **155**, 200–215 (2021).

[10] Chaunzwa, T. L., Hosny, A., Xu, Y., Shafer, A., Diao, N., Lanuti, M., Christiani, D. C., Mak, R. H., and Aerts, H. J., "Deep learning classification of lung cancer histology using ct images," *Scientific reports* **11**(1), 1–12 (2021).

[11] Pittman, K., Sourd, P., Ravnøy, B., Espeland, Ø., Fiksdal, I., Oen, T., Pittman, A., Redmond, K., and Sweetman, J., "Novel method for quantifying salmonid mucous cells," *Journal of fish diseases* **34**(12), 931–936 (2011).

[12] Haddeland, S., Lazado, C. C., Merkin, G. V., Myre, O. J., Okubamichael, M. A., Pedersen, L.-F., and Pittman, K., "Dynamic morphometrics of mucous cells reveal the minimal impact of therapeutic doses of peracetic acid on atlantic salmon gill health," *Aquaculture* **534**, 736315 (2021).

[13] Jayasuriya, N. S., *Gill image analysis : a tool for assessing pathophysiological and morphometric changes in the gill of Atlantic salmon (Salmo salar L.)*, PhD thesis, University of Stirling (2014).

[14] Sveen, L., Timmerhaus, G., Johansen, L.-H., and Ytteborg, E., "Deep neural network analysis-a paradigm shift for histological examination of health and welfare of farmed fish," *Aquaculture* **532**, 736024 (2021).

[15] Maheshwari, S., Pachori, R. B., and Acharya, U. R., "Automated diagnosis of glaucoma using empirical wavelet transform and correntropy features extracted from fundus images," *IEEE journal of biomedical and health informatics* **21**(3), 803–813 (2016).

[16] Kirar, B. S. and Agrawal, D. K., "Computer aided diagnosis of glaucoma using discrete and empirical wavelet transform from fundus images," *IET Image Processing* **13**(1), 73–82 (2019).

[17] Chaudhary, P. K. and Pachori, R. B., "Automatic diagnosis of glaucoma using two-dimensional fourier-bessel series expansion based empirical wavelet transform," *Biomedical Signal Processing and Control* **64**, 102237 (2021).

[18] Reynolds, D. A. et al., "Gaussian mixture models.," *Encyclopedia of biometrics* **741**(659-663) (2009).

[19] Vahadane, A., Peng, T., Sethi, A., Albarqouni, S., Wang, L., Baust, M., Steiger, K., Schlitter, A. M., Esposito, I., and Navab, N., "Structure-preserving color normalization and sparse stain separation for histological images," *IEEE transactions on medical imaging* **35**(8), 1962–1971 (2016).

[20] Krizhevsky, A., Sutskever, I., and Hinton, G. E., "Imagenet classification with deep convolutional neural networks," *Advances in neural information processing systems* **25** (2012).

[21] He, K., Zhang, X., Ren, S., and Sun, J., "Deep residual learning for image recognition," in [*Proceedings of the IEEE conference on computer vision and pattern recognition*], 770–778 (2016).

[22] Howard, A., Sandler, M., Chu, G., Chen, L.-C., Chen, B., Tan, M., Wang, W., Zhu, Y., Pang, R., Vasudevan, V., et al., "Searching for mobilenetv3," in [*Proceedings of the IEEE/CVF International Conference on Computer Vision*], 1314–1324 (2019).

[23] Szegedy, C., Vanhoucke, V., Ioffe, S., Shlens, J., and Wojna, Z., "Rethinking the inception architecture for computer vision," in [*Proceedings of the IEEE conference on computer vision and pattern recognition*], 2818–2826 (2016).

[24] Lee, G., Gommers, R., Waselewski, F., Wohlfahrt, K., and O'Leary, A., "Pywavelets: A python package for wavelet analysis," *Journal of Open Source Software* **4**(36), 1237 (2019).

Appendix 5 - Model Performance metrics.

10.1.1 Model Performance Metrics

The metrics used to evaluate model performance are defined in Equations 10.1, 10.2, 10.3, and 10.4.

$$Accuracy = \frac{TP + TF}{TP + TF + FP + FN} \tag{10.1}$$

$$Precision = \frac{TP}{TP + FP} \tag{10.2}$$

$$Recall = \frac{TP}{TP + FN} \tag{10.3}$$

$$F1 = \frac{2 * Precision * Recall}{Precision + Recall} \tag{10.4}$$