Computer Vision in Wildlife Disease Surveillance

Agnethe Seim Olsen

A thesis submitted for the degree of

Doctor of Philosophy

Cardiff University

August 2025

TABLE OF CONTENTS

List of Figures

Supplementary figures

List of Tables

Supplementary tables

ACKNOWLEDGEMENTS

First and foremost, I want to thank my supervisors, Sarah Perkins and Professor Jo Cable who made me feel valued and capable from our very first meeting. Little did we know that we would not meet again in person until I was quite far into the PhD. I really appreciate your support, both academic and personal. Thank you to Paul Rosin and Chris Jones for your guidance and feedback on everything computer science. Thank you also to Barry Brook, Elise Ringwaldt and Jessie Buettel for allowing me to collaborate with you on your rumpy possums.

I am also grateful to the Ogmore Angling Society who really gave iNaturalist a go and increased the number of brown trout and grayling records in the UK substantially. Many thanks to Bartolomeo Gorgoglione from the Fish Pathobiology and Immunology Laboratory at Michigan State University, the Environment Agency's National Fisheries Laboratory (EA), and Fisheries Management Scotland (FMS) for generously sharing images of infected salmonids. This research would not have been possible without funding from NERC through the GW4 FRESH Doctoral Training Partnership, and I appreciate the camaraderie of every one of my fellow CDT students.

Thank you to all of the students and staff who have endured minutes/hours labelling images of slightly off looking fish. Special thanks to Rhi for the help with pathoPixel and to Neil for months of labelling, writing and coffee support. Thank you to the Lerkins group, who were the only reason I went in to the office most weeks. Thank you to the students and post-docs on the 5th, 6th and 7th floor for moral support, friendship and right on time reminders for lunch breaks. Special thanks to Sarah, Sarah, Gemma, Emma, Dan, Inge, Laura, Guglielmo and Rachel who were there from the start. Diolch to everyone in the organising committee at WEEN and to everyone who attended the conference over the years.

Tusen takk to my friends back in Norway - Marita, Tordis, Renate. The daily updates on goings-on in your lives made me feel included from afar. Takk to my family, both in and out of law. I needed every cat/dog update, visit, holiday, long walk, late dinner and warm hug. I would have grown into my chair without my mother, who met me for weekly strength

workouts over the internet, and my father who reminded us what we were meant to be doing when we were still chatting 30 minutes later.

Thank you to Faraday who was an excellent ice breaker in those early lockdown zoom meetings. And finally, thank you to my husband Joe, who supported me from the moment I decided to leave my stable job to start a PhD in a completely new field at age 30. I could not have done this without you, and this thesis would never have been finished if we had not moved from Bristol to Sheffield, leaving all of those tempting DIY projects behind.

STATEMENT OF CONTRIBUTIONS

The data chapters in this thesis have been compiled as a series of standalone papers or manuscripts, both under review and in preparation for publication. Because each data chapter has been prepared as a standalone paper, there is repetition, especially in the introductions and methods of each chapter.

CHAPTER 1: Introduction

Writing by Agnethe S. Olsen

CHAPTER 2: Seeing Sickness - Computer Vision Precedents and Principles for Wildlife Disease Surveillance

In preparation for submission to an academic journal.

Writing by Agnethe S. Olsen

CHAPTER 3: From Pixels to Pandemics: Quantifying the Potential for Image-Based Wildlife Disease Detection

In preparation for submission to an academic journal.

Agnethe S. Olsen: Conceptualization, Methods, Formal analysis, Data curation, Visualization, Writing - Original Draft, Writing - Review & Editing. **Neil Cook:** Methods, Data curation, Writing – Reviewing & Editing. **Paul L. Rosin**: Methods, Writing - Review & Editing, Funding acquisition. **Christopher B. Jones**: Methods, Writing - Review & Editing, Funding acquisition. **Jo Cable**: Conceptualization, Methods, Writing - Review & Editing, Funding acquisition. **Sarah E. Perkins**: Conceptualization, Methods, Data curation, Writing - Original Draft, Writing - Review & Editing, Funding acquisition.

CHAPTER 4: The extended image: the value of online images for ecological research

In preparation for submission to an academic journal.

Agnethe S. Olsen: Conceptualization, Methods, Formal analysis, Data curation, Visualization, Writing - Original Draft, Writing - Review & Editing. **Neil Cook:** Methods, Data curation, Writing - Original Draft , Writing – Reviewing & Editing. **Paul L. Rosin**: Methods, Writing - Review & Editing, Funding acquisition. **Christopher B. Jones**: Methods, Writing - Review & Editing, Funding acquisition. **Jo Cable**: Conceptualization, Methods, Writing - Original Draft, Writing - Review & Editing, Funding acquisition. **Sarah E. Perkins**: Conceptualization, Methods, Data curation, Writing - Original Draft, Writing - Review & Editing, Funding acquisition.

CHAPTER 5: Computer vision for infectious disease surveillance; *Saprolegnia* spp. in salmonids

A version of this chapter was published in *Ecological Informatics* (Olsen et al. 2026). The thesis version extends that work by adding a cross-evaluation, in which models trained under each dataset strategy are evaluated on identical held-out folds of the Salmo and Oncorhynchus evaluation sets, enabling like-for-like comparisons with genus-specific models; the results, figures and interpretation have been updated accordingly.

Agnethe S. Olsen: Conceptualization, Methods, Software, Formal analysis, Data curation, Visualization, Writing - Original Draft, Writing - Review & Editing. **Paul L. Rosin**: Methods, Writing - Review & Editing, Funding acquisition. **Christopher B. Jones**: Methods, Writing - Review & Editing, Funding acquisition. **Jo Cable**: Conceptualization, Methods, Writing - Original Draft, Writing - Review & Editing, Funding acquisition. **Sarah E. Perkins**: Conceptualization, Methods, Data curation, Writing - Original Draft, Writing - Review & Editing, Funding acquisition.

Acknowledgments

The authors thank Bartolomeo Gorgoglione from the Fish Pathobiology and Immunology Laboratory at Michigan State University, the Environment Agency's National Fisheries Laboratory (EA), and Fisheries Management Scotland (FMS) for generously sharing images used in this study and Neil Cook for image labelling, and Ogmore Angling Association for their input and workshop participation. This work was supported by NERC Standard grant FEC NE/X01049X/1, FRESH - NERC Centre for Doctoral Training in

Freshwater Biosciences and Sustainability NE/R011524/1. We acknowledge the support of the Supercomputing Wales project, which is part-funded by the European Regional Development Fund (ERDF) via Welsh Government.

Statement on inclusion

Our study used a mixture of photographs submitted to online image repositories with no scientific intent (passive citizen science) and some provided directly by fishery science stakeholders. Due to anonymity of image takers, we could not identify and engage with specific individuals. However, to engage with fishers as stakeholders we established a local fisher forum with Ogmore Angling Association, conducting workshops at both the start and conclusion of our study. This group provided valuable feedback on future directions, and we also shared results with the fishery stakeholders who contributed images.

Data availability statement

The code for data processing, model training, and analysis used in this study is publicly available on GitHub at https://github.com/oagn/computer-vision-fish-disease. The metadata for all images sourced from public repositories (iNaturalist, Flickr, GBIF, Wikimedia Commons), including the URLs required to download these images, is available on Zenodo https://doi.org/10.5281/zenodo.15097672 (Olsen et al. 2025). A portion of the 'Saprolegnia spp.' class images (n = 259) was provided directly by stakeholders. These images were shared with us for analysis under specific data sharing agreements with the original contributors and are therefore not publicly available for redistribution.

CHAPTER 6: Deep learning model confidence as a proxy for disease-severity: a case study of rumpwear in possums

In preparation for submission to an academic journal.

Agnethe S. Olsen: Conceptualisation, Methods, Software, Formal analysis, Data curation, Visualisation, Writing - Original Draft, Writing - Review & Editing. **Jessie Buettel:** Data curation, Writing - Review & Editing. **Elise Ringwaldt:** Data curation, Writing - Review & Editing. **Jo Cable:** Writing - Review & Editing, Funding acquisition. **Sarah E. Perkins:** Writing -

Review & Editing, Funding acquisition. **Barry Brook:** Conceptualisation, Methods, Data curation Writing - Review & Editing.

CHAPTER 7: General Discussion and Conclusions

Writing by Agnethe Seim Olsen

THESIS SUMMARY

Wildlife disease surveillance is often limited by the scale, cost, and logistical challenges of traditional methods. This thesis aimed to investigate the potential of using computer vision to classify visible signs of disease in wildlife from digital imagery. In **Chapter 1**, we introduced the challenges of traditional surveillance and the potential for computer vision, while in **Chapter 2** , we provided an overview of the precedents for using computer vision in related fields. In **Chapter 3**, we analysed the WOAH animal disease database and found that over two-thirds (67%) of major infectious diseases present with externally visible signs, and that diseases spread by direct contact were significantly more likely to have them, confirming a broad scope for image-based monitoring. **Chapter 4** then investigated wildlife imagery availability in online repositories, using salmonids as a case study. Our analysis of nearly 70,000 images showed that platforms like iNaturalist are a vast and growing data source, and that a consistent 14-18% of images displayed signs of disease or damage, highlighting an underused resource for health monitoring. In **Chapter 5**, we developed deep learning models to classify *Saprolegnia* spp. infection in salmonids. The best model achieved a high macro-average F1-score (0.930), with performance strongly influenced by dataset composition, as taxonomically focused and balanced datasets yielded the best results. **Chapter 6** used rumpwear in common brushtail possums as a second case study to explore disease severity assessment. We showed a model's output probability for the 'Disease' class served as a robust, well-calibrated proxy for severity, distinguishing between mild and obvious signs, and found that semi-supervised learning provided minimal benefit. Finally, **Chapter 7** discusses how these results show computer vision is a viable tool to complement traditional surveillance, providing a framework to understand disease dynamics and support more timely and effective conservation responses.

CHAPTER 1 :

INTRODUCTION

1.1 Background

1.1.1 Wildlife Disease Monitoring

Effective wildlife disease surveillance represents a critical component of global health security, serving as an early warning system for emerging infectious diseases and zoonotic threats (Daszak et al. 2000; Jones et al. 2008a; Cunningham et al. 2017). Traditional approaches to wildlife disease monitoring rely heavily on resource-intensive field sampling, laboratory diagnostics, and expert pathological assessments (Phelps et al. 2019; Watsa 2020; WOAH 2023). These methods, while effective, face significant limitations in scale, timeliness, and geographical coverage, often exacerbated by funding constraints, lack of harmonised systems, and variable capacity, particularly in remote or resource-limited regions where wildlife health expertise may be scarce (Delgado et al. 2023; Barroso et al. 2024).

The increasing frequency of wildlife disease outbreaks, particularly Emerging Infectious Disease (EID) events, and their potential impacts on biodiversity, ecosystem function, and human health, underscores the urgent need for innovative, scalable, monitoring approaches (Daszak et al. 2000; Jones et al. 2008a; Cunningham et al. 2017). Recent advances in digital technologies, coupled with the proliferation of citizen science initiatives and online biodiversity platforms, offer promising opportunities to complement traditional surveillance methods with novel data streams (August et al. 2020a).

1.1.2 The Potential Target: Visible Signs of Wildlife Disease

A potential avenue for innovative surveillance lies in detecting the externally visible signs of wildlife disease. Veterinary medicine distinguishes between clinical ‘signs’, which are objective indicators noticeable to an observer, and ‘symptoms’, which represent the subjective sensation of illness experienced by the patient. Because animals cannot describe subjective feelings, disease assessment necessarily relies on identifying these observable clinical signs (Constable et al. 2017).

Many wildlife diseases manifest through such externally visible signs, which can include morphological abnormalities (e.g., lesions, tumours, deformities), behavioural changes (e.g.,

lethargy, disorientation), and physiological indicators (e.g., discharge, emaciation) (Ford and Mazzaferro 2012), see

Figure 1.1. Fish species, for instance, exhibit numerous externally visible disease indicators such as skin lesions, fin erosion, exophthalmia (popeye), haemorrhaging, and abnormal growths (Noga 2010). However, the utility of visible signs for monitoring depends heavily on the specific disease and host. Visibility can vary considerably across taxonomic groups, disease causes (aetiologies), and stages of infection. Furthermore, the severity of a disease can alter how subtle its signs are; for example, mild mange may only cause minor fur disruption while severe cases lead to obvious and extensive hair loss. Some conditions, such as prion wasting diseases (e.g., chronic wasting disease), may show no obvious clinical signs until late in infection, making early cases challenging to detect visually (Haley and Hoover 2015). Conversely, diseases such as devil facial tumour disease or white-nose syndrome in bats produce distinctive visual signatures readily documented through photography. Therefore, understanding the spectrum of visually detectable disease manifestations across wildlife taxa is crucial when considering image-based monitoring approaches. The diagnostic value and suitability of targeting visible signs ultimately depend on their specificity, consistency, severity and correlation with underlying disease processes (Stephen and Karesh 2014) — factors which determine their potential for reliable detection in imagery.

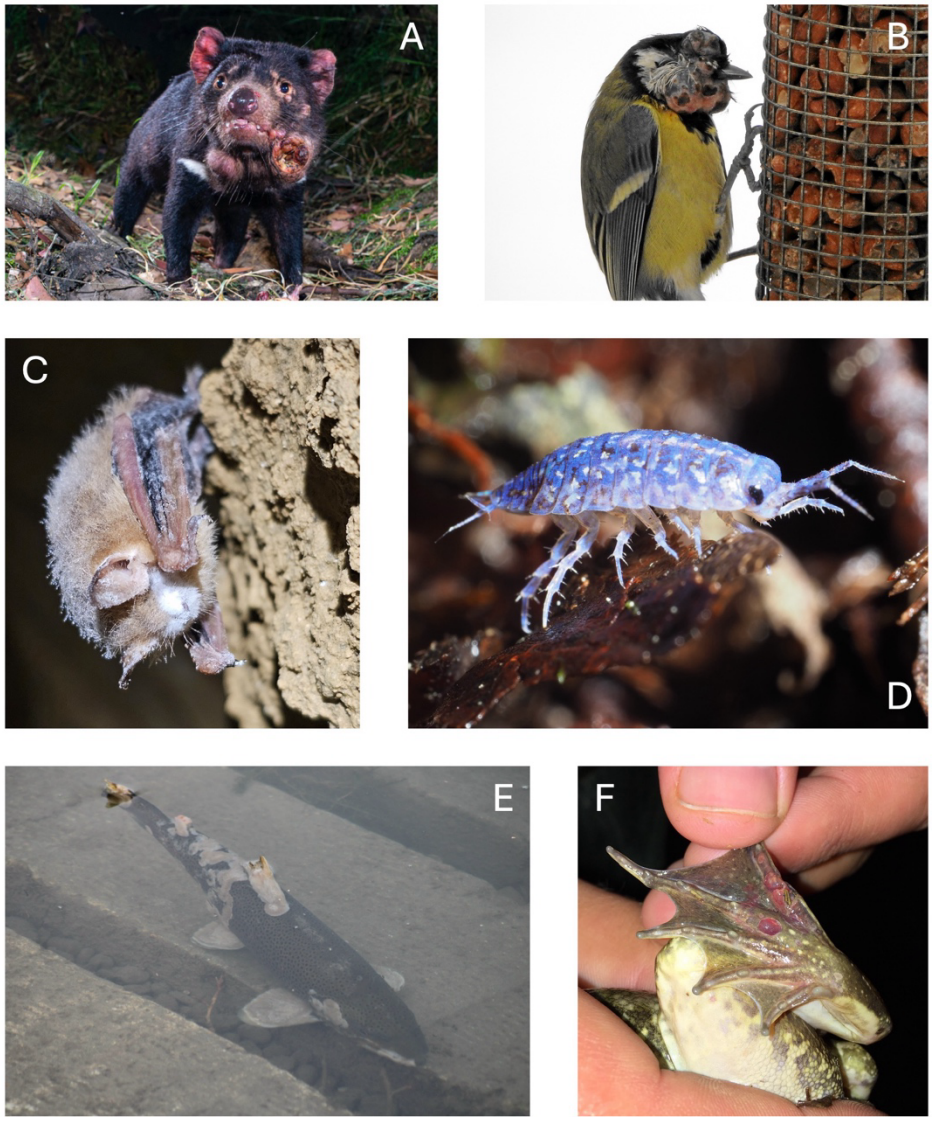


Figure 1.1. Animals presenting with visible signs of disease. A) Tasmanian devil with visible signs of Devil Facial Tumour Disease; B) Blue tit with visible signs of Avian pox; C) Bat with visible signs of White-nose syndrome; D) Woodlouse with visible signs of infection with isopod iridovirus; E) Brown trout with visible signs of infection with *Saprolegnia* spp.; F) Amphibian with visible signs of Amphibian Chytrid. Image Credits: A) Photo 61996475, © Brett Vercoe, some rights reserved (CC BY-NC), uploaded by Brett Vercoe, cropped from the original; B) Photo 180916919, © ingridaltmann, some rights reserved (CC BY-NC); C) Photo 488589421, © Eric C. Maxwell, some rights reserved (CC BY-NC), uploaded by Eric C. Maxwell, cropped from the original; D) Photo 450094682, © Felix Riegel, some rights reserved (CC BY-NC), uploaded by Felix Riegel; E) Photo 337522384, © Sean Cozart, some

rights reserved (CC BY-NC), uploaded by Sean Cozart; F) Photo 7693670, © thesoulflowers, some rights reserved (CC BY-NC), cropped from the original.

1.1.3 Digital Imagery: Abundance, Sources, and the Disease Data Challenge

The digital revolution has generated an unprecedented volume of wildlife imagery, with millions of photographs uploaded annually to diverse online platforms (Nazir and Kaleem 2021; Depauw et al. 2022). These images originate from varied sources, including dedicated citizen science initiatives (e.g., iNaturalist), social media platforms, wildlife photography communities, and institutional repositories (Toivonen et al. 2019). The sheer scale represents a potentially vast resource; for example, analysis conducted for this thesis demonstrates exponential growth in salmonid image submissions to platforms like iNaturalist between 2008 and 2023 (detailed in Chapter 4).

Despite this abundance of wildlife imagery, finding images that document disease within these digital repositories is a significant challenge. This apparent scarcity may stem from several factors. First, the disease itself may be genuinely rare in the wild population. Second, observer biases can play a strong role; societal preferences for charismatic species are well-documented (Troudet et al. 2017), and it is plausible that observers also preferentially photograph or share images of healthy-looking animals. Furthermore, limited awareness of disease indicators among non-specialists and platform designs not focused on disease reporting likely exacerbate this scarcity. Nevertheless, these digital repositories likely contain valuable, largely untapped information for wildlife health monitoring, as platforms not explicitly focused on disease may still inadvertently capture visual evidence of health abnormalities. However, harnessing this potential requires overcoming significant inherent challenges. Extracting reliable ecological insights from such vast, non-systematically generated online datasets necessitates careful consideration of data biases, validation requirements, and interpretation difficulties (Jarić et al. 2020). Applied to wildlife health, this manifests as the core challenge of efficiently identifying potentially rare disease-relevant images, validating the presence of visible signs (often from limited visual information and non-expert sources), and analysing these findings amidst data heterogeneity (Jarić et al. 2020), a task further complicated by the fact that most visible

signs are not pathognomonic and could indicate multiple conditions (Stephen and Karesh 2014). Addressing this complex task necessitates powerful automated tools capable of processing images at scale.

1.2 Computer Vision: A Tool for Scalable Image Analysis

Addressing the challenge of analysing vast repositories of wildlife imagery for rare disease signs requires powerful automated tools. Computer vision is the field that develops computational methods to extract information from digital images and video. Some computer vision approaches use rule-based image processing, while others use artificial intelligence methods. This thesis focuses on artificial intelligence approaches to computer vision, mainly machine learning. These methods learn patterns from data rather than relying on hand-designed rules. A particular focus is deep learning, which uses neural networks with many layers to learn useful image features directly from pixel data (LeCun et al. 2015). See Figure 1.2 for a visual representation of these related fields.

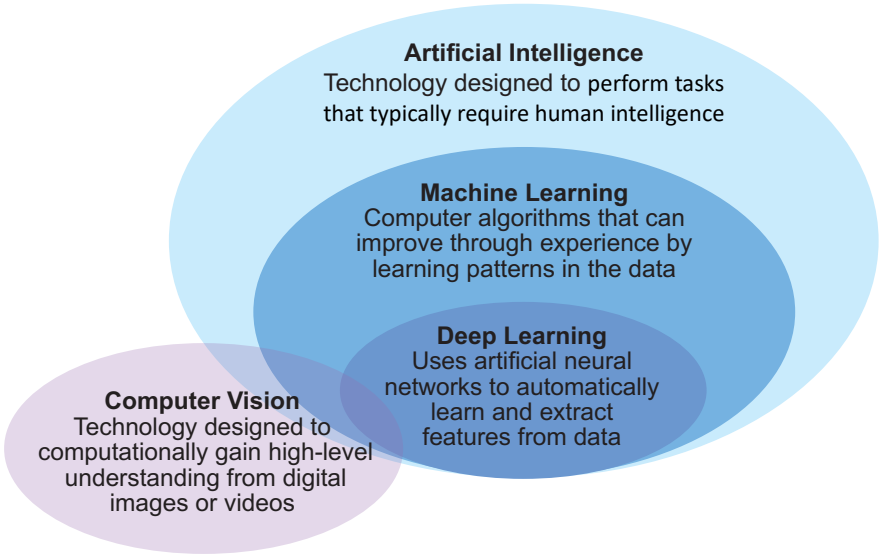


Figure 1.2. Conceptual relationship between Artificial Intelligence, Machine Learning, Deep Learning, and Computer Vision.

These technologies, especially deep learning approaches like Convolutional Neural Networks (a type of neural network particularly adept at processing grid-like data such as images), have revolutionised automated image analysis across numerous domains. Their success in core tasks such as object detection, image classification, and semantic segmentation makes them highly promising for application to wildlife disease monitoring through automated analysis of visual data.

Employing computer vision for wildlife health assessment presents several key advantages over manual inspection. These include the scalability required to process massive image datasets often generated by camera traps and citizen science efforts (Green et al. 2020), the potential for near real-time analysis of incoming data streams (Arshad et al. 2020), and the ability, demonstrated prominently in medical imaging, to potentially detect subtle visual patterns indicative of early-stage disease that might elude human observers (Esteva et al. 2017). Furthermore, these systems can theoretically improve over time through continuous learning as more validated training data becomes available. However, applying these technologies effectively to wildlife disease detection faces significant hurdles. Key challenges include severe class imbalance issues stemming from the relative scarcity of disease images compared to healthy ones, a well-documented problem in ecological datasets (Cunha et al. 2023), the critical need for substantial amounts of high-quality, expertly labelled training data, which is laborious and costly to obtain for wildlife disease imagery (Green et al. 2020; Cunha et al. 2023), the presence of numerous confounding factors inherent in field photography such as variable illumination, pose, distance, and background complexity (Cunha et al. 2023), and the inherent complexity, similar to challenges in medical imaging, of reliably distinguishing subtle pathological conditions from normal biological variation or environmental artefacts (Wei et al. 2022). Overcoming these challenges is central to the research presented in this thesis.

1.3 Research Aim and Questions

The **aim** of this thesis is to investigate the potential and challenges of using computer vision technologies to classify visible signs of disease in wildlife through the analysis of digital imagery. To achieve this aim, the following specific **research questions** are addressed:

1: Which wildlife diseases, particularly those subject to existing monitoring efforts, present with externally visible signs suitable for detection via computer vision?

2: What is the availability, quality, and ecological information content of wildlife imagery, particularly for fish species (salmonids), accessible from online digital repositories, and what is the prevalence of visible disease signs within these data?

3: How effectively can deep learning algorithms be developed and trained to classify specific, visible signs of disease in wildlife images, using *Saprolegnia* spp. in salmonids and 'rumpwear' in common brushtail possums (*Trichosurus vulpecula*) as case studies, and how does dataset composition influence model performance?

4: Can the outputs of computer vision classification models serve as a proxy for disease severity assessment, and can semi-supervised learning approaches help mitigate labelled data limitations in this context?

1.4 Research Design

To address the research aim and questions, this thesis employs a multi-faceted research design integrating literature synthesis, database analysis, and empirical computer vision case studies.

Initially, a foundation is established through reviews of relevant literature, exploring the application of computer vision for disease detection in related fields and outlining key methodological considerations (**Chapter 2**). Building on this, a systematic analysis of the WOAH animal diseases databases is conducted to identify and characterise diseases presenting with externally visible signs, thereby assessing the scope for computer vision-based monitoring (addressing Question 1 in **Chapter 3**). This involves collating disease metadata and classifying visible signs according to defined criteria.

The practical feasibility is then investigated through quantitative assessment of data availability, focusing on the characteristics and potential utility of wildlife imagery sourced

from online digital repositories (such as Flickr and iNaturalist), exemplified by a detailed case study on salmonids (addressing Question 2 in **Chapter 4**).

Finally, the core potential of the approach is evaluated empirically through the development, training, and assessment of deep learning-based computer vision pipelines. Two distinct case studies are presented: one focusing on the classification of *Saprolegnia* spp. in salmonids using different dataset compositions (addressing Question 3 in **Chapter 5**), and another exploring disease severity assessment and semi-supervised learning techniques for possum rumpwear (addressing Question 4 in **Chapter 6**). Statistical analyses, including General Linear Models and goodness-of-fit tests where appropriate, are used to analyse results from the database and case study investigations.

This sequential design allows the research to move from defining the theoretical potential and scope (Chapters 2 and 3), through understanding data limitations (Chapter 4), to demonstrating practical implementation and methodological refinement using specific wildlife disease examples (Chapters 5 and 6).

1.5 Thesis Structure

This thesis is structured into seven chapters, including this introduction, to systematically address the research aim and questions. The four data chapters (Chapters 3-6) are presented as self-contained studies, each written with the intention of publication in a peer-reviewed journal.

Chapter 1: Introduction (this chapter) introduces the research context, defines the problem of wildlife disease surveillance, outlines the potential of computer vision, establishes the research aim and specific questions, describes the research design, and provides this roadmap.

Chapter 2: Seeing Sickness: Computer Vision Precedents and Principles for Wildlife Disease Surveillance provides essential context by reviewing the application of computer vision for disease detection in related fields (human medicine, livestock, plant pathology), critically discusses the unique challenges and opportunities presented by wildlife disease ecology,

and outlines foundational methodologies for data handling and computer vision pipeline development relevant to this thesis.

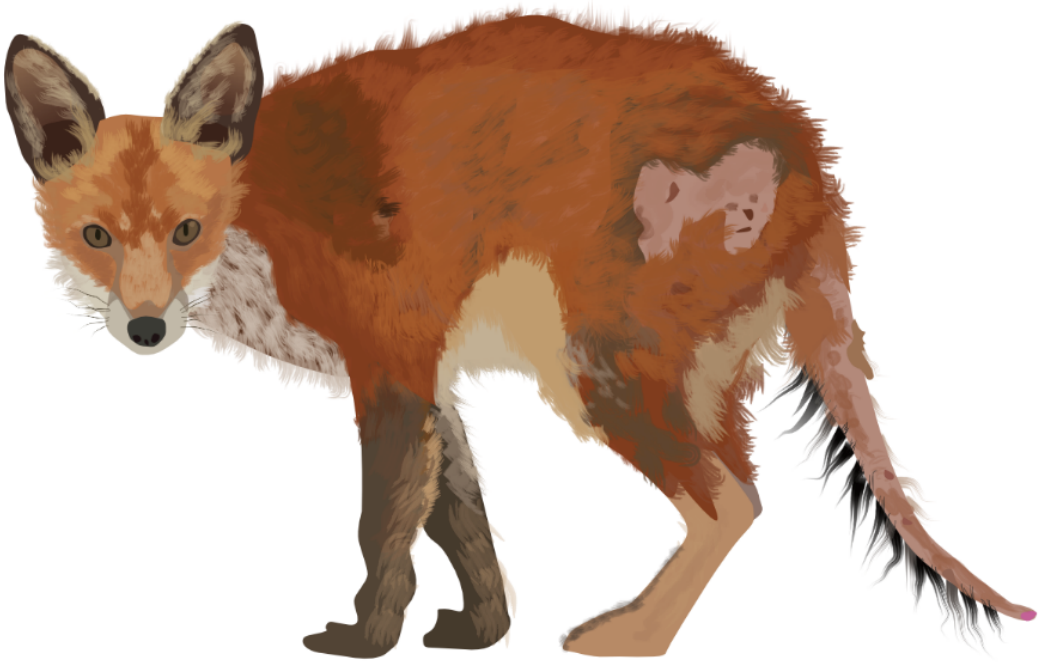
Chapter 3: From Pixels to Pandemics: Quantifying the Potential for Image-Based Wildlife Disease Detection conducts a systematic analysis of the WOAH animal disease database to identify and characterise wildlife diseases that present with externally visible signs, thereby assessing their suitability as potential targets for computer vision-based detection approaches.

Chapter 4: The Extended Image: the Value of Online Images for Ecological Research investigates the availability, characteristics, and potential utility of wildlife imagery from online digital repositories (e.g., iNaturalist, Flickr) for disease monitoring, focusing specifically on salmonid fishes as a case study to understand data landscapes and limitations.

Chapter 5: Computer Vision for Infectious Disease Surveillance; *Saprolegnia* spp. in Salmonids presents the development, training, and evaluation of a deep learning pipeline designed to classify visible signs of *Saprolegnia* spp. infection in salmonids, exploring the influence of dataset composition and taxonomic specificity on model performance.

Chapter 6: Deep Learning Model Confidence as a Proxy for Disease-Severity: a Case Study of Rumpwear in Possums explores the application of computer vision beyond simple detection, investigating its potential for assessing disease severity using possum rump wear as a case study and examining the utility of semi-supervised learning techniques to mitigate labelled data limitations.

Chapter 7: General Discussion and Conclusions synthesises the key findings from the preceding chapters, discusses the overall implications and limitations of using computer vision for wildlife disease surveillance based on this research, and proposes directions for future work in this emerging field.



CHAPTER 2 :

SEEING SICKNESS - COMPUTER VISION

PRECEDENTS AND PRINCIPLES FOR WILDLIFE

DISEASE SURVEILLANCE

Abstract

Computer vision can improve wildlife disease surveillance by allowing for the non-invasive, large-scale analysis of images from sources such as camera traps and citizen science. However, applying computer vision methods that have worked well in controlled settings like human medicine and agriculture to wildlife in natural environments presents specific challenges. This review provides a practical introduction for ecologists, synthesising the opportunities, challenges, and key steps for developing and evaluating computer vision systems for this purpose. We outline a workflow from data acquisition and labelling to model development and validation, using sarcoptic mange in red foxes as a recurring example. While the main opportunity lies in using large, opportunistic image archives to augment traditional surveillance, the most significant challenges are data centric. These include the scarcity of disease images, high variability in image quality, and the difficulty of obtaining reliable ground-truth labels for training and validation. We conclude that effective implementation requires a tailored, end-to-end workflow addressing these specific data hurdles.

2.1 Introduction

Computer vision (CV), powered by advances in machine learning and the increasing availability of visual data, offers transformative potential for ecological research and wildlife monitoring (Weinstein 2018; Tuia et al. 2022). Beyond established applications in species identification and behavioural analysis, a burgeoning area of interest lies in leveraging CV for non-invasive wildlife disease surveillance, aiming to augment traditional epidemiological methods that can be resource-intensive and challenging to implement at scale in wild populations (Delgado et al. 2023; Barroso et al. 2024). Detecting visible signs of disease through automated image analysis could provide valuable insights into disease prevalence, distribution, and host impacts.

Translating CV techniques successfully to the complexities of wildlife disease, however, requires careful consideration. While applications in related fields like human medicine (Esteva et al. 2021), livestock health (Fuentes et al. 2022) and plant disease (Ferentinos 2018) offer valuable precedents, the unique challenges of working with wild animals in uncontrolled environments necessitate a tailored approach. Therefore, before delving into

the specific wildlife disease applications explored later in this thesis, this chapter aims to provide a foundational overview essential for developing and evaluating CV systems in this context. To achieve this, the chapter is structured as follows: we first highlight the existing gaps in wildlife disease surveillance and introduce the rationale for image-based monitoring (Section 2.2). Next, we explain the core foundational concepts of computer vision, machine learning, and deep learning relevant to understanding the subsequent discussions (Section 2.3). We then briefly survey the established uses of CV within the broader field of animal ecology (Section 2.4) before examining precedents for disease detection in the related domains of human medicine, plant pathology, and livestock/aquaculture health (Section 2.5). The core of the chapter (Section 2.6) then critically analyses the specific opportunities and inherent challenges of applying computer vision to wildlife disease surveillance, integrating discussion of the key practical considerations involved in implementation; from data acquisition and labelling to model development and evaluation. To ground these technical discussions in a practical context, we will use sarcoptic mange in red foxes (*Vulpes vulpes*) as a recurring case example throughout the chapter to illustrate key concepts and challenges. Finally, this chapter concludes with a summary and forward look, outlining a conceptual framework for future work in this emerging field (Section 2.7).

2.2 The Wildlife Disease Surveillance Gap and the Potential Role of Image-Based Monitoring

Despite the recognised importance of monitoring animal diseases, global surveillance efforts remain inconsistent. While frameworks exist for reporting notifiable diseases, particularly in livestock (e.g., through World Organisation for Animal Health, WOAH), implementation is often patchy, and systematic surveillance for diseases in wildlife populations is generally considered *ad hoc* (Phelps et al. 2019; Watsa et al. 2020). Gathering diagnostic samples from wildlife presents significant logistical challenges, contributing to these surveillance gaps (Watsa 2020; Delgado et al. 2023; Barroso et al. 2024). Consequently, the initial detection of wildlife disease outbreaks often relies heavily on opportunistic observations by the public – including hunters, anglers, wildlife enthusiasts, and photographers – who report sightings of sick or deceased animals. Several significant epizootics were first identified through such observations of externally visible signs. For example, an epidemic of

trichomonosis in British finches was identified following unsolicited reports from the public of sick and dead birds at garden feeders (Robinson et al. 2010), while the spread of mycoplasmal conjunctivitis in North American house finches was monitored through a large-scale citizen-science survey of feeder birds (Dhondt et al. 1998).

The critical role of visible signs of disease highlights a potential pathway for enhancing surveillance. If disease signs are visually apparent to observers, they can, in principle, be captured in photographs or videos. This motivates the exploration of image-based monitoring, leveraging the increasing abundance of digital imagery alongside CV tools, as a non-invasive approach to help bridge existing gaps in wildlife disease surveillance.

2.3 Computer Vision: Foundational Concepts

Ecological research increasingly uses image and video data for non-invasive monitoring, generating vast datasets that require efficient analysis methods (Tuia et al. 2022; Pollock et al. 2025). CV is the field of computer science focused on enabling machines to "see" and interpret visual information from the world, much like human vision. It offers powerful tools to automate tasks like identifying species, counting individuals, or, pertinent to this thesis, detecting signs of disease within these large image collections (Weinstein 2018; Pollock et al. 2025). While applying these methods effectively often benefits from collaboration with computer scientists (Weinstein 2018; Ditria et al. 2020; Vidal et al. 2021), understanding the core concepts is essential for ecologists seeking to use these techniques. Here, we provide a primer on key CV concepts, particularly those relevant to image-based disease detection.

2.3.1 Machine learning concepts for computer vision

At its heart, much of modern CV relies on Machine Learning (ML), a branch of Artificial Intelligence (AI) where systems learn to perform tasks from data without being explicitly programmed for every step (Braga-Neto 2024). Instead of defining rigid rules, ML models identify patterns within data to make predictions or decisions. Learning techniques are broadly categorised as supervised or unsupervised. In supervised learning, the most common approach for tasks like disease classification, the model is trained on labelled training data; images that have been manually annotated with the correct answer or class (e.g., explicitly labelled as, for example, "mange present" or "mange absent"). During an iterative process called training, the model learns to associate visual patterns, often called

features; informative image characteristics based on aspects like colour, texture, edges, or corners - with the correct labels by comparing its predictions to the provided ground truth labels. In contrast, unsupervised learning algorithms attempt to discover inherent structures or patterns within unlabelled data. While evaluating the intrinsic success of purely unsupervised pattern discovery (like clustering) can be less direct than measuring accuracy against known labels (Valletta et al. 2017), techniques that learn representations from unlabelled data (such as self-supervised learning; He et al. 2020) have become extremely powerful. These learned representations can then be effectively adapted for specific tasks, like classification, using smaller amounts of labelled data. However, given the primary goal of identifying and classifying specific, known disease conditions in this thesis, this chapter focuses mainly on supervised learning approaches, while acknowledging the growing importance of methods that leverage unlabelled data.

2.3.2 From feature engineering to deep learning

Historically, supervised machine learning in CV often involved a two-step process (Figure 2.1a). First, a human expert, the "feature engineer," would carefully select and design algorithms known as feature descriptors. These algorithms automatically detect relevant visual patterns (features) and encode them, often into a series of numbers that could be used to differentiate one feature type from another. These hand-crafted features were then fed into a separate classifier algorithm (e.g., a Support Vector Machine) which learned to distinguish between classes based on those predefined features (Szeliski 2022). This traditional approach required significant domain expertise and laborious fine-tuning to identify the most effective features for a given task (O'Mahony et al. 2020).

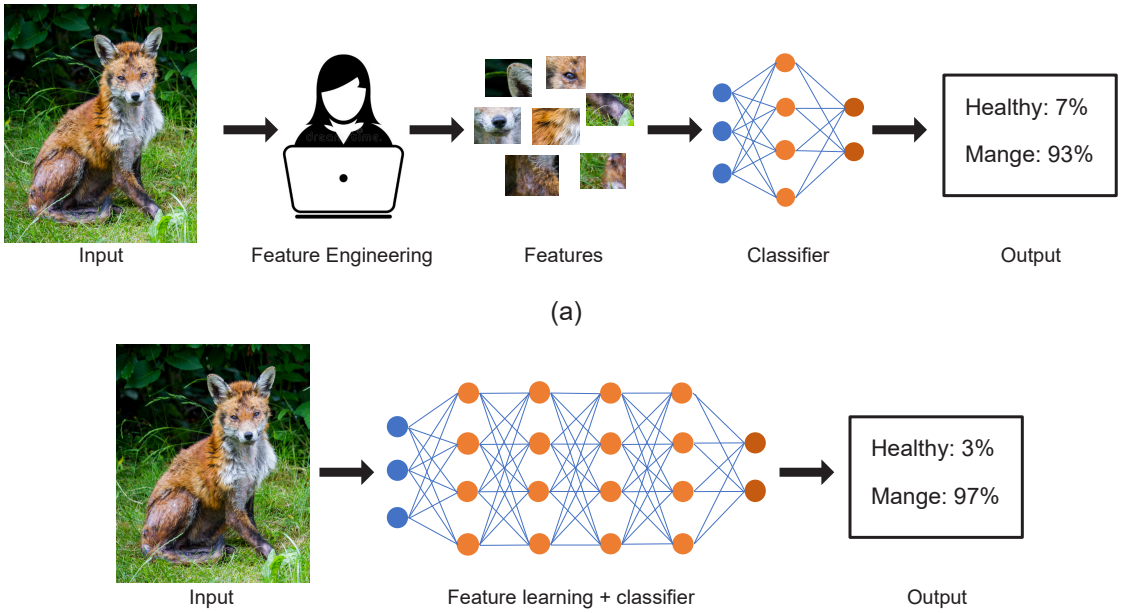


Figure 2.1 Comparison of traditional and deep learning computer vision classification workflows. (a) Traditional computer vision classification workflow. Features hand-crafted by a feature engineer are extracted from the input data and used as input for a classifier. This requires detailed knowledge of feature descriptors. (b) Deep learning classification workflow. The deep learning model performs feature learning and classification directly from the input images and their corresponding labels in an end-to-end manner. The output percentages are included as examples to illustrate potential model outputs. Image copyright: Mangey fox in my garden, © Paul Williams, some rights reserved (CC BY-NC), uploaded by Paul Williams, cropped from the original image.

A major shift occurred with the rise of Deep Learning (DL), a subfield of ML (LeCun et al. 2015). DL models, particularly those based on Artificial Neural Networks (ANNs), enable "end-to-end" learning (Figure 2.1b) (LeCun et al. 2015). Instead of relying on hand-crafted features, DL algorithms automatically learn the most relevant descriptive features directly

from the labelled input data during training (O'Mahony et al. 2020). ANNs are inspired by the structure of biological brains, composed of interconnected processing units called artificial neurons, typically organised in layers (LeCun et al. 2015). Information flows from an input layer, through one or more 'hidden' layers, to an output layer. Connections between neurons have associated weights that are adjusted during training, strengthening or weakening signals based on how well the network performs the target task (LeCun et al. 2015).

For image analysis, Convolutional Neural Networks (CNNs) have become the dominant DL architecture (LeCun et al. 2015). CNNs incorporate specialised 'convolutional' layers that are good at detecting spatial hierarchies of patterns within images. Early layers might detect simple features like edges or corners, while deeper layers combine these to recognise more complex shapes, textures, and eventually objects (e.g., parts of an animal, specific lesion types or severity of disease). The development of the 'AlexNet' CNN (Krizhevsky et al. 2012), which significantly outperformed previous methods on a major image recognition challenge (ImageNet), spurred rapid advancements. Since then, numerous CNN architectures have been developed, such as VGG (Simonyan and Zisserman 2014), ResNet (He et al. 2016), Inception (Szegedy et al. 2016), and EfficientNet (Tan and Le 2019), offering different trade-offs in performance, size, and computational cost (Khan et al. 2020). More recently, architectures based on the Transformer model, originally developed for natural language processing, have also shown strong performance on image analysis tasks (Dosovitskiy et al. 2021).

2.3.3 Computer vision tasks and annotation requirements

These DL models can be trained for various CV tasks relevant to disease ecology, differing primarily in the type of information extracted and the required labelling (Figure 2.2) (Szeliski 2022). The simplest task is image classification, where the entire image is assigned a single label (e.g., "mange", "healthy"). This requires only image-level labels (Figure 2.2a), but a potential pitfall is that the model might learn to associate the label with irrelevant background features if there are systematic differences between classes (e.g., if most mange photos are taken in urban areas and healthy ones in rural areas) (Miao et al. 2019). A more advanced task is object detection, which involves not only classifying objects within an image but also localising them, typically by drawing a bounding box around each instance

(Figure 2.2b). This requires more detailed labelling but allows for detecting multiple animals or lesions in one image and can help mitigate background dependence issues. Common architectures used for this include Faster R-CNN (Ren et al. 2015) and YOLO (Redmon et al. 2015; Bochkovskiy et al. 2020). For the most detailed spatial information, semantic Segmentation assigns a class label to every pixel in the image, effectively creating a mask outlining the object (Figure 2.2c) or affected area (Figure 2.2d). This requires intensive pixel-level annotation but can be valuable for quantifying the extent of lesions, for example. While models like Mask R-CNN (He et al. 2017) are popular architectures for this task, the manual annotation burden often makes a human-in-the-loop process more practical. In such workflows, a model generates a draft segmentation that a human expert can quickly approve or refine. This interactive approach not only accelerates the creation of new training data but also facilitates continuous model improvement by using the corrections to retrain the model, a strategy also known as active learning.

Figure 2.2. Different levels of image annotation for training Convolutional Neural Networks. (a) Image-level label for classification (e.g., **'**Fox with mange**'**). (b) Object detection using bounding boxes to localise and classify regions of interest. (c) Semantic segmentation using polygons (or masks) to classify each pixel belonging to the region of interest 'Fox with mange'. (d) Semantic segmentation using polygons (or masks) to classify each pixel belonging to the region of interest 'Visible signs of mange'. Image copyright: Mangey fox in

my garden, © Paul Williams, some rights reserved (CC BY-NC), uploaded by Paul Williams, cropped and edited from the original image.

2.3.4 Evaluating model performance

Evaluating the performance of these different tasks requires specific metrics tailored to what is being measured. For image classification, where the task is a binary decision, labelling a whole image as either 'fox with mange' or 'healthy fox', common metrics derive from a confusion matrix comparing predictions to these ground truth labels. This matrix counts: True Positives (TP) where the model correctly identifies a fox with mange; True Negatives (TN) where it correctly identifies a healthy fox; False Positives (FP) where it wrongly labels a healthy fox as having mange; and False Negatives (FN) where it misses mange in a fox that actually has it. While overall Accuracy ((TP+TN)/Total) provides a general sense of correctness, it can be misleading if, as is common in disease studies, one class (e.g., healthy foxes) is much more frequent than the other. Therefore, Precision (TP/(TP+FP)) is crucial: it tells us, out of all the foxes that the model claimed had mange, the proportion that did, indicating the rate of false alarms. Recall (or Sensitivity, TP/(TP+FN)) is equally important, measuring the proportion of all foxes that genuinely had mange that the model successfully identified. The F1-score, a harmonic mean of Precision and Recall, offers a single metric balancing the risk of false alarms against the risk of missing cases. Specificity (TN/(TN+FP)) measures how well the model correctly identifies the healthy foxes. Crucially, from a disease surveillance perspective, the relative importance of metrics like Precision and Recall depends on the ecological or management context; a high rate of false negatives (low Recall) might mean missing critical early warnings, while excessive false positives (low Precision) could lead to unnecessary and costly interventions.

For object detection, where the goal might be to specifically localise the visible signs of mange by drawing a bounding box around affected areas (rather than just the whole fox), evaluation must consider both the classification correctness (is it mange?) and the localisation accuracy (is the box placed correctly?). The overlap between the predicted bounding box around the mange signs and the ground truth box is typically quantified using Intersection over Union (IoU) calculated as the area of overlap between the predicted and ground truth boxes divided by the total area encompassed by both boxes. Standard metrics

like mean Average Precision (mAP) then summarise the detection performance across different confidence thresholds, integrating both how well the mange signs were classified and how accurately they were localised via the bounding box, judged against a specific IoU threshold.

Finally, for semantic segmentation, which aims to create a precise pixel-level mask outlining only the fox, or the mange-affected skin areas on the fox, evaluation focuses on the accuracy of this mask. Metrics commonly used are again IoU (sometimes called the Jaccard Index in segmentation) or the Dice Coefficient (closely related to the F1-score but calculated at the pixel level). These quantify how well the pixels predicted by the model as being part of a mange lesion overlap with the true lesion pixels defined in the ground truth mask. Understanding the specific information provided by these different metrics is crucial for interpreting model performance and selecting or developing models appropriate for specific ecological or epidemiological questions related to wildlife disease.

2.3.5 Practical approaches and emerging methods

Training deep CNNs from scratch typically require vast amounts of labelled data and substantial computational resources (Zhuang et al. 2020). Transfer Learning offers a practical solution. It uses models pre-trained on large, general-purpose datasets (like ImageNet, containing millions of diverse images). By adapting these pre-trained models and fine-tuning only some layers on a smaller, task-specific dataset (e.g., wildlife disease images), researchers can often achieve high performance with significantly less data and computational effort, essentially transferring the general visual knowledge learned from the large dataset to the new problem (Yosinski et al. 2014).

Meta Learning, or "learning to learn," represents another advanced area of machine learning (Hospedales et al. 2022). Instead of learning to classify specific objects directly, these methods learn a process for comparing images or learning new classes quickly from very few examples (few-shot learning). Techniques like metric learning (e.g., Siamese networks, triplet loss) learn similarity functions, which can be useful for tasks like identifying new or unseen anomalies (Hospedales et al. 2022). This ability to handle previously unseen classes relates to the concept of Open Set Recognition, where a model must not only classify known categories but also identify inputs that do not belong to any of the known classes,

which is relevant when encountering unexpected diseases or conditions in surveillance data (Barcina-Blanco et al. 2024).

Beyond these learning paradigms, recent advances in Vision-Language Models (VLMs) offer new ways of interacting with and analysing image data (Radford et al. 2021; Liu et al. 2023b). These models, often built upon Large Language Model architectures, learn joint representations of images and text, enabling tasks like zero-shot classification, visual question answering, and text-to-image retrieval. The ability to query large image datasets using natural language descriptions holds significant potential for ecological research, allowing scientists to search for specific visual concepts (e.g., behaviours, interactions, visible signs) that may not be captured by standard metadata or classification labels. However, evaluating the effectiveness of these models for complex, domain-specific queries remains an active area of research, with benchmarks specifically designed for ecological applications highlighting current limitations, particularly with expert-level concepts and fine-grained visual details (Vendrow et al. 2024).

Understanding these fundamental concepts, which include supervised learning, the shift from feature engineering to deep learning with CNNs and Transformers, the different CV tasks and their labelling requirements, practical techniques like transfer learning, and emerging capabilities like meta-learning and vision-language interaction, provides the necessary foundation for evaluating the application of CV in related fields (Section 2.5) and understanding the specific opportunities and challenges within wildlife disease ecology (Section 2.6).

2.4 Computer Vision Applications in Animal Ecology

The application of computer vision in ecological research involving animals has grown exponentially, particularly over the past decade, driven by the increasing availability of digital visual data and the need to automate its processing and analysis (Jarić et al. 2020; Pollock et al. 2025). A Scopus search for publications combining terms related to computer vision and animals shows a marked increase over time (Figure 2.3a). In contrast, the number of publications linking computer vision, animals, and disease is much lower, although it has also increased in recent years (Figure 2.3a). To account for background growth in computer vision, Figure 2.3b shows these trends normalised by the annual number of “computer

vision" publications. In an offset regression with a negative binomial robustness check, the rate of "computer vision AND (wildlife OR animal*)" increased by about 3.8% per year, while "computer vision AND (wildlife OR animal*) AND disease" increased by about 10.7% per year. An interaction model indicated that the disease-related subset increased faster than the broader animals subset (p =). Broad search terms such as "animal*" may include laboratory animal studies rather than wildlife research, which may inflate the apparent volume of general "CV Animals" work. Nonetheless, the trend indicates increasing use of these methods in animal-related research, alongside a growing but smaller body of work applying computer vision to disease.

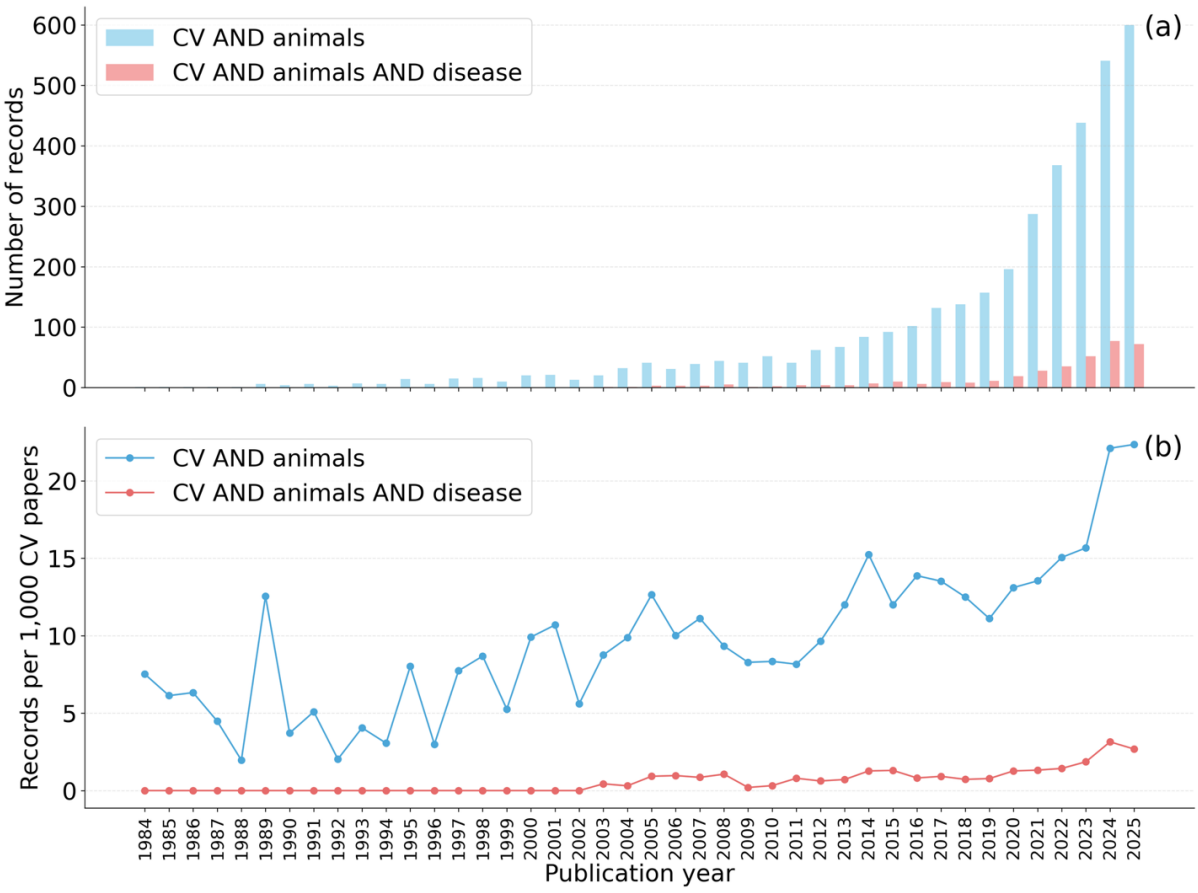


Figure 2.3. Scopus publication trends. (a) The annual number of publications (articles, reviews, conference papers and data papers) returned for the search terms '"computer vision" AND (wildlife OR animal*)' and '"computer vision" AND (wildlife OR animal*) AND disease' on Scopus from 1984- 2025. (b) The same trends expressed as records per 1,000 background 'computer vision' publications per year. The background rate is the annual

number of publications returned for the search term ‘computer vision’ using the same document type filters and time period.

Within ecology, CV powered by deep learning is now routinely leveraged for a wide range of monitoring tasks previously reliant on manual effort (Tuia et al. 2022; Pollock et al. 2025). Key applications include automating the description, counting, and classification of animal species from images and video, crucial for biodiversity assessments and species management (Weinstein 2018; Willi et al. 2019; Norouzzadeh et al. 2021). Beyond species identification, CV is also widely employed for analysing animal behaviour, such as social interactions, quantifying activity patterns, facilitating the re-identification of individuals based on unique markings, and tracking animal movements (Robie et al. 2017; Schneider et al. 2019; Ravoor and Sudarshan 2020; Ratnayake et al. 2021).

The visual data used is highly diverse, spanning scales from satellite and drone imagery down to microscopy, and originating from sources as varied as smartphones used in controlled settings or by citizen scientists (Spiesman et al. 2021), camera traps (Norouzzadeh et al. 2021), and specialised sensors operating beyond the human visual spectrum, such as infrared cameras (Dunbar et al. 2009). The adoption of deep learning algorithms, particularly Convolutional Neural Networks (CNNs) discussed in Section 2.3.2, has significantly advanced these applications, often achieving classification accuracies similar to, or even exceeding, human expert performance for tasks like species identification (Swanson et al. 2015; Gómez-Villa et al. 2016; Torney et al. 2019; Ditria et al. 2020).

Given this demonstrated success of CV in extracting diverse ecological information from various image types, yet the comparatively limited focus on disease applications evident in the literature trends (Figure 2.3), its potential application specifically for detecting signs of disease warrants detailed investigation, building upon techniques established in other health-related domains.

2.5 Computer Vision for Disease Detection in Related Domains

While the application of CV specifically to wildlife disease surveillance is nascent, considerable progress has been made in using these techniques for disease detection and diagnosis in the related fields of human medicine, plant disease, and livestock/aquaculture health. Examining these precedents provides valuable insights into established methodologies, potential capabilities, and challenges relevant to wildlife applications.

In human medicine, CV is now well-integrated into many areas of diagnostics, analysing a wide range of imaging modalities (Esteva et al. 2021). Numerous AI-based medical technologies using CV have received regulatory approval (e.g., at least 29 by the US FDA between 2012-2020), aiding image reconstruction, analysis, and diagnosis (Benjamens et al. 2020). These tools are applied across fields including radiology (analysing X-rays, CT, MRI) (Rajkomar et al. 2017; Weston et al. 2019; Jain et al. 2021), cardiology (Wehbe et al. 2023), and ophthalmology (analysing retinal scans) (Kucur et al. 2018; Panda et al. 2018; Hemelings et al. 2020). Particularly relevant to monitoring visible signs in wildlife, CV systems analysing standard clinical photographs in dermatology have achieved diagnostic abilities comparable to, or even exceeding, those of human specialists for tasks like classifying skin lesions (Esteva et al. 2017; Haenssle et al. 2018; Liu et al. 2020). Beyond diagnosis, CV assists in diverse clinical tasks such as population screening, predicting patient outcomes, segmenting pathological structures (at organ or cellular levels), and monitoring disease progression over time (Esteva et al. 2021). The success in medicine is often driven by the availability of large, curated datasets and the relatively standardised nature of many medical images (Esteva et al. 2021). However, the fundamental pattern recognition capabilities demonstrated, especially in analysing photographic images for conditions like skin diseases, highlight the potential for similar approaches in animals, while acknowledging the significant differences in image quality and consistency expected from wildlife data.

Similarly, automated disease identification using CV has gained significant traction in agriculture for crop protection. Numerous studies have applied techniques, often deep learning, to classify diseases based on images of plant leaves, stems, or fruits (Mohanty et al. 2016; Sladojevic et al. 2016a; Amara et al. 2017; Gui et al. 2021). High classification accuracies, sometimes exceeding 99% under specific conditions, have been reported for

identifying multiple diseases across various plant species, demonstrating the potential for automated systems in field or controlled environments (Ferentinos 2018; Bhagwat and Dandawate 2021). This success underscores the utility of CV for detecting visual disease signs (like spots, lesions, wilting, colour changes) in biological organisms using standard imagery, albeit typically under more controlled conditions than wildlife monitoring.

Image-based disease detection is comparatively less developed for monitoring the health of mobile animals like livestock or wildlife, but research is growing. In livestock farming, efforts include using CV for automated lameness detection in cattle based on gait analysis (Barney et al. 2023; Myint et al. 2024), monitoring respiratory diseases through behavioural or thermal changes (Jorquera-Chavez et al. 2020; Wu et al. 2023), or identifying skin conditions (Rony et al. 2021) and external parasites (Barbedo et al. 2017). In aquaculture, disease-related applications often focus on controlled environments, targeting issues like external parasite detection (e.g., sea lice on salmon) or identifying visible lesions indicative of bacterial or fungal infection (Liu et al. 2023a). While progress has been made, some studies in this area have faced limitations related to dataset size, image quality, or methodological validation (see Chapter 5, this thesis; Liu et al. 2023).

Commercial aquaculture has also developed imaging systems for fish health and welfare monitoring, particularly in salmon farming where external parasites such as sea lice are routinely assessed. A number of commercial suppliers now market camera-based and computer-vision systems for salmon aquaculture (e.g., Aquabyte and OptoScale), with applications that commonly include automated sea-lice assessment, fish sizing, and welfare/condition scoring. These systems typically use fixed in-cage or underwater cameras to capture images or video of fish as they pass and then apply automated image analysis to produce operational metrics. Such outputs are also relevant to ecological and behavioural questions. However, much of the underlying imagery, derived data streams, and implementation detail from commercial deployments are proprietary or otherwise not publicly accessible, which limits independent evaluation, reproducibility, and reuse in academic research.

While still limited, some studies have demonstrated the potential of CV for detecting visible disease signs specifically in wildlife contexts. For instance, photographic analysis has been

explored for monitoring the progression of Devil Facial Tumour Disease (DFTD) based on lesion characteristics (Nurçin et al. 2024), and researchers have applied image analysis techniques to assess skin conditions in marine mammals like bottlenose dolphins (*Tursiops erebennus*) (Murphy et al. 2025). Similarly, computer vision techniques have been used to classify images of brushtail possums based on the presence and severity of 'rumpwear' a condition involving fur loss (Chapter 6, this thesis). Work detailed in this thesis (Chapter 5) also demonstrates the application of deep learning to classify visible signs of *Saprolegnia* spp. infection in salmonids from photographic images. These examples, though fewer than in human medicine or plant disease, serve as important precedents validating the exploration of CV for detecting visually distinct signs of disease or abnormal conditions in diverse animal taxa, including wildlife.

Synthesising these applications shows that computer vision can classify visible signs of disease from images, sometimes reaching expert-level performance with large, curated datasets from controlled settings. The main CV tasks used, which we described in Section 2.3.3, include classification, detection, and segmentation. While supervised learning is common, research in these fields is also exploring advanced techniques like few-shot learning and open-set recognition to address data limitations (Singh et al. 2021; Dong et al. 2024). These methods are relevant to the data scarcity and potential for new conditions found in wildlife surveillance. However, translating this success to wildlife disease monitoring is not straightforward. The main difficulties come from the uncontrolled nature of wildlife imagery and the challenge of obtaining enough accurately labelled images (Green et al. 2020; Cunha et al. 2023). Therefore, while these precedents are encouraging, we cannot directly transfer the methods. Developing effective tools requires tailored approaches to address these data challenges, which we explore in the following section.

2.6 Applying Computer Vision to Wildlife Disease Surveillance: Opportunities, Challenges, and Practical Considerations

Having established the potential of CV in ecological research (Section 2.3.4) and its demonstrated utility in disease detection within different fields (Section 0), this section focuses on the application of these technologies to wildlife disease surveillance. It examines the significant opportunities presented by CV in this context, addresses the substantial

challenges inherent in working with wildlife populations and data, and outlines the key practical considerations necessary for successful implementation.

2.6.1 Opportunities in CV-Based Surveillance

The integration of CV into wildlife disease surveillance offers several compelling advantages compared to traditional methodologies, which often rely on invasive sampling or opportunistic carcass discovery (Watsa 2020; Delgado et al. 2023), limiting scalability and timeliness. A primary benefit is the potential for enhanced non-invasive monitoring. CV allows for the assessment of visible disease signs from digital imagery (photographs and videos) without requiring the physical capture or handling of animals. This significantly reduces potential stress to wildlife and alleviates the considerable logistical challenges and costs associated with capturing and sampling free-ranging, often elusive, species (Burton et al. 2015). Examples include identifying skin lesions in cetaceans from photographic surveys (Murphy et al. 2025) and monitoring visible signs of poor health or external lesions in terrestrial mammals using camera-trap imagery (Muneza et al. 2019; Murray et al. 2021).

CV also enables greater scalability in monitoring efforts. Ecological studies using camera traps or citizen science platforms can generate enormous volumes of visual data; for instance, the Snapshot Serengeti project amassed over 1.2 million image sets within three years (Swanson et al. 2015). Manual review of such datasets is prohibitively time-consuming. CV provides the necessary tools to automate the processing of these large image collections, facilitating tasks like species identification and potentially flagging images exhibiting disease indicators, thereby transforming data analysis from a bottleneck into a more tractable process (Norouzzadeh et al. 2018; Green et al. 2020).

Automation, in turn, allows for an expanded scope of surveillance. Monitoring can be extended across larger geographical areas and sustained over longer durations than is often feasible using traditional field-based methods alone. Furthermore, CV techniques offer the exciting potential for application to retrospectively analyse archived image collections, potentially uncovering historical disease patterns or tracking pathogen spread over timeframes previously inaccessible. The frequent availability of metadata associated with digital images, such as time, date, and location information (from EXIF data or camera deployment records), adds significant value (Biggs et al. 2009; Daume 2016). Integrating

these spatio-temporal data points with CV-derived health classifications facilitates epidemiological analyses, including mapping disease distributions, tracking potential spread, and investigating associations with environmental variables (Toivonen et al. 2019).

Finally, there is potential for earlier detection of disease outbreaks or changes in prevalence. Continuous or wide-scale monitoring using automated image analysis might identify subtle visual signs across a population before they become apparent through more limited traditional surveillance approaches. Citizen science initiatives, where numerous observers contribute images, represent a particularly promising avenue for broad-scale detection if participants can be effectively guided in recognising and reporting relevant signs (Scott et al. 2020).

2.6.2 Challenges and Practical Steps

Despite the significant potential, the effective application of CV to wildlife disease surveillance necessitates addressing several substantial challenges related to data, methodology, and implementation.

Getting the Right Ingredients: Data Acquisition and Curation

A fundamental difficulty lies in obtaining sufficient high-quality image data depicting the target disease(s). While digital wildlife imagery is abundant overall, images clearly showing specific disease signs are often scarce (Green et al. 2020). This rarity may stem from the low prevalence of certain diseases, the subtle nature of early-stage signs, or potential observer bias against photographing unhealthy-appearing animals. This frequently results in highly imbalanced datasets, where healthy individuals vastly outnumber diseased ones, posing a significant challenge for training unbiased machine learning models.

Overcoming this data scarcity necessitates a strategic and often multi-faceted approach to image sourcing. Several potential avenues exist, each presenting distinct opportunities and limitations. Large camera trap archives such as Wildlife Insights (wildlifeinsights.org), LILA (lila.science) and Agouti (Casaer et al. 2019), for instance, offer extensive systematic data, though isolating relevant disease images might require targeted searches or specialised analytical techniques (Murray et al. 2021). Citizen science platforms such as iNaturalist provide large quantities of geo-referenced, opportunistic data (Van Horn et al. 2018).Their

value for ecological research, including health assessment, is increasingly recognised (Vendrow et al. 2024). These data are particularly relevant for monitoring species that interact frequently with the public or recreational users, such as fish observed by anglers or boaters. However, using these platforms for disease monitoring may depend on project designs that encourage relevant reporting(Scott et al. 2020). While social media yields immense image quantities (Durso et al. 2021; Edwards et al. 2021), researchers must grapple with significant challenges related to data noise, inherent biases, and ethical usage (Morcatty et al. 2024). In contrast, collaborations with veterinary or rehabilitation facilities can provide high-quality images of confirmed cases, offering valuable ground truth, albeit potentially lacking natural ecological context. Given these varied characteristics, integrating data from multiple sources is often the most pragmatic strategy. Regardless of the source(s), even when employing transfer learning (Section 2.3.5) to reduce data requirements (Yosinski et al. 2014), securing several hundred labelled examples per class (including 'healthy') is generally advisable for robust model development (Christin et al. 2019). Furthermore, careful management and utilisation of associated metadata, such as location and time, is crucial throughout the acquisition and curation process to maximise the potential for subsequent ecological analysis (Toivonen et al. 2019).

Seeing Clearly: Data Quality and Labelling

Beyond the initial acquisition, the inherent quality and subsequent labelling of wildlife imagery present further significant hurdles for computer vision applications. Images captured in natural environments typically exhibit extreme variability in factors such as lighting, animal pose, distance, occlusion, and background complexity (Figure 2.4). This often results in a low signal-to-noise ratio, where subtle disease signs can be difficult to discern amidst visual clutter (Lürig et al. 2021; Cunha et al. 2023). These challenges can be particularly acute in aquatic environments, where factors such as water turbidity, surface reflections, rapid subject movement, and the inherent difficulties of underwater imaging further reduce image clarity and the visibility of subtle signs. Compounding this issue, visual diagnosis from images alone can be inherently ambiguous; differentiating visual signs of disease from natural variation or injuries can prove challenging, even for experts. Such ambiguity inevitably complicates the creation of reliable ground truth labels necessary for training supervised models (Murray et al. 2021).

Addressing these combined challenges requires meticulous attention to data quality control and the establishment of robust labelling procedures. Paramount among these is the development of clear, objective definitions for disease categories coupled with consistent labelling protocols applied across the dataset. The specific type of annotation required must also align carefully with the intended analytical goal: simple image-level labels suffice for basic classification tasks, whereas object detection necessitates bounding boxes to localise signs, and semantic segmentation demands detailed pixel-level masks for precise spatial delineation (Figure 2.2). Selecting appropriate labelling software (such as CVAT or Zooniverse) can facilitate this process. Critically, given the diagnostic difficulty inherent in identifying wildlife diseases visually, ensuring high label quality typically necessitates significant domain expertise (veterinary or ecological) (Murray et al. 2021). This contrasts sharply with tasks like species identification, where non-expert citizen scientists can often achieve high accuracy (Swanson et al. 2015). To further improve reliability and quantify uncertainty, best practice involves employing multiple annotators for each image (or at least a subset) and assessing consistency through metrics like Inter-Annotator Agreement (IAA) (Artstein 2017; Ditria et al. 2020; Palmer et al. 2021). While comparison against a 'gold standard' set with clinically confirmed diagnoses represents the ideal for label validation, this approach is unfortunately rarely feasible in free-ranging wildlife contexts, underscoring the importance of rigorous protocol design and expert involvement.

Figure 2.4. Examples of photographs of red foxes (*Vulpes vulpes*) with clinical signs of sarcoptic mange, sourced from the citizen science platform iNaturalist. The panels illustrate common challenges and characteristics, such as: (A) a blurry, low-light camera trap image; (B) a clear, high-quality photograph suitable for detailed assessment; (C) a subject that is distant in an image that has been edited; (D) a subject that is both blurry and partially occluded by foliage; (E) an image clearly showing clinical signs of mange along the flank and tail; and (F) a subject at a significant distance from the camera. Image Credits: (A) Photo 401908233, © Nova Patch, CC BY-SA; (B) Photo 248359548, © Lori A Owenby, CC BY-NC; (C) Photo 61922852, © Juan C. Espinosa, CC BY; (D) Photo 333198287, © Erin O'Connor, CC BY-NC; (E) Photo 360827452, © Lyla R. Meader, CC BY-NC; (F) Photo 223123668, © Cole Wolf, CC BY. All images were uploaded by their respective owners to iNaturalist.

Building the Engine: Model Development and Methodological Complexity

Developing effective computer vision models for wildlife disease classification or detection presents considerable methodological challenges beyond data acquisition and labelling. The visual signs themselves can be subtle, non-specific, mimic normal variation, or change significantly with disease stage, making automated detection intrinsically difficult (Wei et al. 2022). A further complication arises from the risk of models learning spurious correlations; they might associate disease labels with irrelevant background elements present in the training data rather than the relevant disease features, thereby severely limiting their ability to generalise to new images or environments (Miao et al. 2019). Even advanced Vision-Language Models (VLMs), despite their potential for flexible data interaction, currently face hurdles in this domain, often struggling with the fine-grained visual distinctions required for accurate disease identification and lacking comprehension of specialised terminology used in expert queries (Vendrow et al. 2024). Looming over these technical difficulties is the fundamental challenge of rigorously validating model outputs against confirmed ('gold standard') disease status in free-ranging wildlife populations. This is a significant, often challenging, obstacle given the inability to capture and clinically assess most individuals observed remotely.

Mitigating these complex challenges requires a thoughtful and rigorous approach to model development and training. Careful workflow design is crucial, involving the selection of

appropriate computer vision tasks (classification, detection, or segmentation) tailored to the specific research questions and potentially employing multi-stage pipelines, for instance, using an initial general animal detector like MegaDetector (Beery et al. 2019) before applying a specialised disease classifier. To leverage existing knowledge when specialised datasets are small or scarce, transfer learning (Yosinski et al. 2014) is a standard approach where models pre-trained on large, general datasets are adapted for the specific wildlife disease task. Furthermore, advanced learning paradigms offer promising avenues to capitalise on potentially abundant unlabelled imagery alongside limited labelled examples. Active learning strategies, for example, can intelligently prioritise the most informative images for expert annotation, maximising model improvement while minimising labelling effort. Techniques like semi-supervised learning explicitly incorporate unlabelled data into the training process to improve model generalisation, while self-supervised learning methods can first learn rich visual representations from unlabelled data alone, which are then fine-tuned using the available labels (He et al. 2020). Such approaches, along with few-shot learning strategies designed to learn effectively from minimal labelled examples (Section 2.3.5), are particularly pertinent given the typical constraints of wildlife disease datasets. Essential pre-processing steps, such as image resizing and pixel value normalisation, ensure data consistency for model input. Furthermore, data augmentation, applying random transformations like flips, rotations, and brightness adjustments to training images, is critically important for enhancing model robustness against the inherent variability in wildlife photographs (Shorten and Khoshgoftaar 2019). Emerging generative AI techniques also offer a potential, though still developing, avenue for synthesising data for particularly rare classes (Rafiq et al. 2025). Finally, the model training process itself demands careful management. This includes using distinct training and validation data splits, ideally incorporating spatial or temporal separation to rigorously test generalisation capabilities (Beery et al. 2018; Tabak et al. 2019; Schneider et al. 2020), which helps monitor learning progress and prevent the common problem of overfitting where the model performs well on training data but poorly on unseen data. This entire workflow, from data acquisition and annotation through to model evaluation and interpretation, is outlined in Figure 2.5.

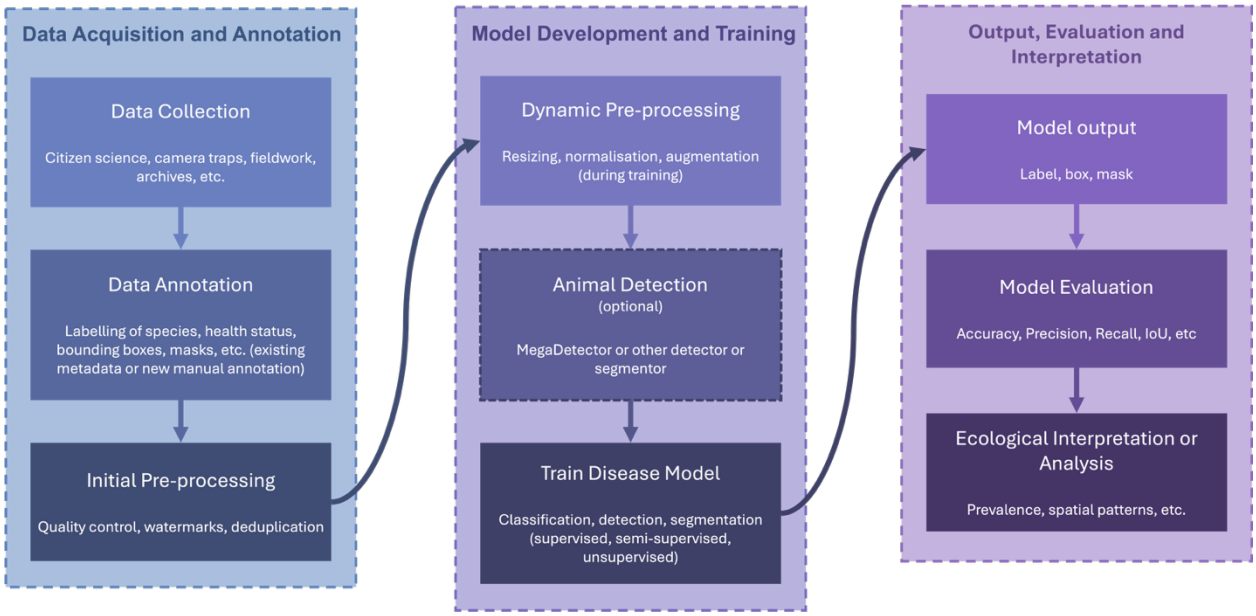


Figure 2.5. Schematic overview of the computer vision pipeline for wildlife disease detection from online images. Data are collected and annotated, undergo initial quality control, and are dynamically pre-processed during model training. Optional animal detection can be applied before disease classification or segmentation. Model outputs are evaluated and interpreted in an ecological context.

Checking the Results: Evaluation, Interpretation and Deployment

Once a model is trained, evaluating its performance presents its own set of challenges, particularly within ecological contexts. Relying solely on overall accuracy can be highly misleading, especially given the class imbalance typically encountered in disease datasets where healthy individuals vastly outnumber diseased ones. Furthermore, demonstrating that a model achieves high statistical performance on standard metrics (as defined in Section 2.3.4) does not automatically guarantee that its outputs provide meaningful ecological insights or accurately reflect real-world disease dynamics.

Consequently, a nuanced and comprehensive evaluation strategy is essential. This involves selecting performance metrics appropriate for the specific task and sensitive to data characteristics, particularly the class imbalance inherent in many disease studies (Blair et al.

2024). Beyond calculating these core metrics, evaluation should include comparing model performance to human expert benchmarks where feasible, providing valuable context for its capabilities (Esteva et al. 2017). Robustness and generalisability must also be rigorously assessed, ideally by testing the model on data collected under different conditions or from different locations than those represented in the training set (Beery et al. 2018; Ditria et al. 2020). Ideally, evaluation should extend beyond standard computer vision metrics to assess whether model performance translates into the accurate estimation of relevant downstream ecological variables, such as prevalence estimates, as this link is not always direct (Pantazis et al. 2024). To gain further confidence in the model reasoning, model interpretation techniques, such as saliency heat maps which highlight image regions influencing predictions, can help verify that the model is focusing on relevant pathological features rather than spurious background cues (Selvaraju et al. 2020; see Figure 2.6). Only through thorough evaluation can the suitability of a model for deployment and interpretation be confidently determined.

Figure 2.6. Illustration of a saliency-style heatmap for post-hoc model interpretation. The image on the left shows a red fox (Vulpes vulpes) with visible signs consistent with sarcoptic mange. The panel on the right shows a heatmap overlay created manually in Affinity Designer (using a semi-transparent colour layer) to illustrate how saliency visualisations are

commonly presented. Warmer colours indicate regions of higher apparent importance. This figure is illustrative only and is not an output from any model used in this thesis.

Putting it Together: Resource and Ethical Considerations

Finally, successfully implementing computer vision for wildlife disease surveillance involves navigating significant resource constraints and ethical responsibilities. Deep learning workflows typically demand considerable computational power for training and inference and require specialised technical expertise in data science and programming. These requirements can present substantial barriers for ecology research groups operating with limited budgets or lacking dedicated technical staff (Ditria et al. 2020; Vidal et al. 2021). Concurrently, the handling of wildlife health data necessitates careful consideration of ethical issues, particularly regarding data sensitivity and potential impacts on conservation efforts or public perception.

Addressing these practical hurdles often requires interdisciplinary collaboration. Partnerships between ecologists and computer or data scientists can bridge expertise gaps and are frequently essential for project success (Weinstein 2018). Effective project planning must explicitly account for computational requirements, potentially exploring options like cloud computing resources or shared infrastructure. Importantly, all stages of the research, from data acquisition through to analysis and dissemination, must adhere to ethical protocols (Wilkinson et al. 2016). This includes ensuring appropriate permissions for data collection, implementing secure data storage, and carefully managing the sharing of potentially sensitive information, such as precise locations of vulnerable species or graphic images depicting specific disease conditions (Morcatty et al. 2024). Careful attention to both resource management and ethical conduct is fundamental to the responsible application of computer vision in this sensitive field.

2.7 Summary and Forward Look

This review has provided a foundational overview of the potential for using computer vision in wildlife disease surveillance. We have seen that while precedents from human medicine,

agriculture, and livestock health demonstrate the considerable capabilities of CV for analysing visual signs of disease (Section 0), the successful application to wildlife requires careful navigation of unique challenges (Section 2.6). Key opportunities include the potential for non-invasive monitoring at large scales, leveraging the vast amounts of available digital imagery (Section 2.6.1). However, significant hurdles exist concerning data availability and quality, the inherent ambiguity of visual diagnosis in variable field conditions, methodological complexities in model development, and practical considerations regarding resources and ethics (Section 2.6.2).

Successfully applying CV effectively to wildlife disease monitoring, therefore, requires more than adopting off-the-shelf tools. It necessitates a nuanced, integrated approach that addresses these specific challenges. Success hinges on developing targeted data curation strategies, selecting appropriate model architectures and workflows, employing robust validation methods that consider ecological relevance beyond model accuracy, and fostering strong interdisciplinary collaboration. Future advancements, perhaps leveraging the capabilities of Vision-Language Models, might even open new avenues for detecting novel or unexpected disease signs that have not been explicitly trained for. The groundwork laid in this chapter provides the necessary context for the subsequent chapters of this thesis, which explore these elements through the development and evaluation of CV tools designed for specific wildlife disease case studies.

CHAPTER 3 :

From Pixels to Pandemics: Quantifying the Potential for Image-Based Infectious Disease Surveillance of Animals

Authors: Agnethe Olsen[1], Chris B. Jones[2], Paul Rosin[2], Neil Cook[1], Jo Cable[1], Sarah E. Perkins[1]

Affiliations:

[1]School of Biosciences and Water Research Institute, Cardiff University, CF10 3AX, UK.

[2]School of Computer Science and Informatics, Cardiff University, CF24 4AG, UK.

ABSTRACT

Animal disease surveillance is dependent on early detection for an optimal response to both emerging and known pathogens. For wildlife, diseases are often reported *ad hoc* presenting a challenge for both conservation and public health. Images acquired via smartphones, camera traps, and open-access websites, however, present an opportunity for non-invasive surveillance of diseases with visible signs that could supplement existing, sparse knowledge.

Here, we quantify the visible signs associated with animal diseases listed on the World Organisation for Animal Health (WOAH) animal diseases portal. From a database of 204 unique infectious diseases, we found that 137 (67.2%) have at least one visible sign. These signs were most frequently associated with viral pathogens and mammalian hosts, with 'lesion' being the most commonly recorded sign. Diseases that spread by direct contact were significantly more likely to have visible signs than those transmitted indirectly or by vectors. To assess the feasibility of image-based surveillance, we interrogated 'Flickr', an open access image sharing website and found public images for 74 of the 137 diseases with visible signs.

While we show the potential for image-based disease detection in a range of host-pathogen systems, challenges remain in acquiring sufficient number and diversity of images and associated metadata for comprehensive spatial and temporal surveillance. Actively acquiring images via a camera or sensor networks, or from stakeholder citizen science groups, are likely the most fruitful means to acquire sufficient spatiotemporal coverage for surveillance. Thereafter, sufficient images could provide an opportunity to train a computer vision model for surveillance. These data could enable near real-time feedback and generate

novel spatiotemporal data for determining, together with location data, when, where, and why disease occurs in animals.

3.1 Introduction

Epizootics cause significant animal welfare issues, lead to population declines, and in wildlife can threaten species viability (McCallum 2008). Widespread disease and high mortality are obvious conservation concerns, especially when coupled with pervasive environmental stressors such as climate change and pollution (Smith et al. 2009). Up to 80% of the range of the northern long-eared bat (*Myotis septentrionalis*) in North America, for example, is at risk from white-nose syndrome, with widespread mortality occurring (Cheng et al. 2021). Similarly, devil facial tumour disease (DFTD) has spread rapidly throughout the Tasmanian devil (*Sarcophilus harrisii*) population in Australia, causing up to 90% mortality (Jones et al. 2008b). Infectious diseases that spill over into humans (zoonoses) can yield devastating societal health and socio-economic impacts (Morens et al. 2004; Karesh et al. 2012). Over 70% of emerging human infectious diseases originate from wildlife (Jones et al. 2008a) which provides strong motivation for animal disease surveillance. The impact of animal disease on conservation, food security, animal welfare and human health therefore renders surveillance vital (Dobson et al. 2006; Muneza et al. 2019; Savary et al. 2019). The economic incentive for zoonotic surveillance is obvious: the SARS-CoV-2 pandemic resulted in estimated economic losses of $14 trillion in the US alone (Walmsley et al. 2021). Despite considerable potential for harm and these clear incentives for surveillance, our understanding of where and when animal diseases occur, especially within wild species, is sporadic and incomplete (Watsa 2020), a gap highlighted in the most recent international guidelines on the topic (WOAH and IUCN 2024).

One promising avenue for closing this surveillance gap is through non-invasive, image-based methods. Many infectious diseases produce externally visible signs, which can be captured in photographs.

3.1.1 Disease Surveillance of Animals

While global infrastructure for animal disease surveillance exists, it is primarily focused on livestock and commercially important species, with comparatively limited surveillance in wildlife (WOAH and IUCN 2024). Member countries of the World Organisation for Animal

Health (WOAH) have a duty to report 'listed' terrestrial and aquatic animal diseases; those considered of international concern by WOAH. Reports typically stem from clinical surveillance (veterinary observations), syndromic surveillance (systematic analysis of health data), ante- and post-mortem inspections (including in slaughterhouses/abattoirs) and surveillance of sentinel animals (WOAH 2023). Coordinated pathogen screening is clearly an important tool to prevent epizootics/pandemics, but depends on having animals 'in-hand' and available for inspection, and while this may be relatively straightforward for livestock or domestic animals, 'point of care diagnostics' remains a significant challenge for wildlife (Bora et al. 2022; WOAH and IUCN 2024).

Surveillance and disease detection in wildlife are typically *ad hoc* (Phelps et al. 2019; Watsa 2020) and often reports originate from the general public. Individuals that interact with animals, such as hunters, anglers and wildlife enthusiasts, are likely to report disease outbreaks or sick animals due to their understanding of what is unusual (Mörner et al. 2002). Take, for example, the emergence of white nose syndrome, where the pathogen itself is often visible on the face and was first noticed by recreational cavers in bat hibernacula (Hoyt et al. 2021). Similarly, DFTD causes large lesions, typically on the face, and was first detected by a wildlife photographer (Hawkins et al. 2006). These diseases illustrate that where signs are visible to the human eye then it is theoretically possible to use image-based disease detection for surveillance (Figure 3.1).

Figure 3.1. Photographs from online sources showing wildlife species infected with diseases with visible signs. (A) Red fox, *Vulpes vulpes*, with visible signs of mange. (B) Green sea turtle, *Chelonia mydas*, with visible signs of myxomatosis. (C) Common raven, *Corvus corax*, with visible signs of avian pox. (D) Rabbit, *Oryctolagus cuniculus*, with visible signs of myxomatosis. (E) Limosa harlequin frog, *Atelopus limosus*, with visible signs of chytridiomycosis. (F) Coho salmon, *Oncorhynchus kisutch*, with visible signs of *Saprolegnia* spp. See Supplementary Material, Chapter 3, for photograph attributions.

3.1.2 Image-Based Infectious Disease Detection in Animals

Non-invasive imaging offers a promising avenue for wildlife disease surveillance. Infrared thermography (IRT), for example, has previously been used as a non-invasive method to measure spatial variation in body surface temperature as a proxy for host health (Schilling et

al. 2022). This has been used to successfully detect disease, notably sarcoptic mange (*Sarcoptes scabiei*) that results in heterogenous fur loss and so is associated with heat loss (Escobar et al. 2022). Such imaging, however, requires specialist equipment, is labour intensive, and disease detection performance is comparable to visual observation (Arenas et al. 2002). In contrast, standard photographs offer a simple, versatile method for collecting what the latest international guidelines term "non-biological samples" (WOAH and IUCN 2024). Smartphone images, which often contain valuable spatiotemporal metadata, and images from camera traps or public photo-sharing platforms like Flickr, provide a rich source of data (Fox et al. 2020; Terry et al. 2020; Edwards et al. 2021). Engaging citizen scientists to contribute such images is a form of participatory surveillance that can help overcome coverage gaps and detect spatial hotspots, as demonstrated with mange in red foxes (*Vulpes vulpes*) (Scott et al. 2020). Yet, despite available infrastructure, image-based disease surveillance in animals remains rare (Schilling et al. 2022). While manual screening is viable at small scales, computer vision offers a scalable, cost-effective solution. Overall, this approach provides a rapid, non-invasive method to help close surveillance gaps.

3.1.3 Computer Vision for Infectious Disease Detection

Although many ecological studies now use computer vision for species classification (Norouzzadeh et al. 2021) and to record species distributions (Beery et al. 2021), few have applied it to animal disease detection in the wild. Outside ecology, however, computer vision is already used operationally for health and welfare monitoring in managed animal systems. In Precision Livestock Farming (Berckmans 2017), commercial vision systems such as HerdVision provide automated body condition and mobility scoring (HerdVision 2021). Similar camera-based approaches are also used in aquaculture and fisheries, for example to support sea-lice monitoring in salmon production (Tidal 2025) and AI-assisted analysis of catch imagery for sustainable management (Skirrow 2024). In research settings, the relatively small number of existing studies nonetheless shows that computer vision can screen large image datasets for health-related signals, motivating further development for wildlife disease surveillance (Christin et al. 2019; Jarić et al. 2020; Poulin et al. 2021). For example, Park et al. (2007) reported 80-90% identification accuracy for scuticociliates and *Trichodina* spp. in the olive flounder (*Paralichthys olivaceus*). More recently, Olsen et al. in press developed a computer vision pipeline to classify saprolegniasis-like infections in wild

salmonids, achieving high performance using images compiled from citizen science platforms. In contrast, plant pathology has more broadly adopted computer vision. Automated systems for *in situ* crop disease detection have achieved high accuracy (Mohanty et al. 2016; Sladojevic et al. 2016b; Amara et al. 2017; Boulent et al. 2019; Bischoff et al. 2021), with some studies reporting up to 99% accuracy for multiple diseases across a range of plant species (Fuentes et al. 2017; Ferentinos 2018; Bhagwat and Dandawate 2021). Medical image analysis is even more advanced, spanning multiple modalities and demonstrating diagnostic performance rivalling or surpassing human experts (Esteva et al. 2017; Nishio et al. 2018; Jain et al. 2021). These successes are largely attributed to the abundance of high-quality, well-annotated training images available in these domains. The primary challenge for animal disease surveillance, therefore, lies in acquiring sufficient images to develop and train similar models. This study represents a first step in that direction by quantifying which diseases might be suitable for this approach and assessing the current availability of images, providing a roadmap for future work in this area.

In this study, we quantify this potential by systematically assessing animal diseases of international concern for visible signs that could be used for image-based detection. We focus on infectious diseases because their capacity for transmission makes them a primary target for early detection and large-scale surveillance. Specific aims are to: (1) identify and quantify important animal diseases that have visible signs and therefore potential use in image-based disease detection; (2) identify general characteristics of diseases and their hosts by testing the association between visible signs and metadata on pathogen taxonomy, host taxonomy, transmission mode, host context (wildlife or livestock), and zoonotic potential; and (3) assess the availability of open-access digital imagery to evaluate the feasibility of image-based surveillance for identified diseases.

3.2 Methods

3.2.1 Animal Disease Database Construction

As no single comprehensive database of all animal infectious diseases exists, we first compiled and curated a foundational list of infectious diseases from the World Organisation for Animal Health animal diseases portal (WOAH, 2020). We selected this source for two reasons: first, these diseases are subject to global attention and are typically well-

documented, allowing for the reliable extraction of metadata on clinical signs. Second, these diseases were more likely to have publicly available image data, relevant to our aim of assessing the potential for image-based surveillance.

The list of diseases sourced from the WOAH portal (n=206) underwent a three-step curation process to establish a consistent naming convention. First, we excluded three non-infectious agents (e.g., botulism). Second, we disaggregated broad entries into their constituent diseases (e.g., resolving the single entry for 'Zoonoses transmissible from non-human primates' into 45 specific diseases listed within the WOAH Terrestrial Manual (WOAH 2023), of which 23 did not already have entries for mammals in the database), yielding a total of 250 diseases. Third, we standardised disease names to create a single canonical label for each disease (e.g., records for 'Equine Influenza' and 'Equine Influenza, in wildlife' were both mapped to the standardised entity 'Equine Influenza'). This curation process yielded a Master Disease List of 204 unique diseases.

For each of the 204 diseases in our Master Disease List, we systematically extracted detailed information on hosts, pathogens, clinical signs, and transmission to create a Detailed Disease-Host Database. Our primary source was the WOAH portal, which we supplemented with the Merck Veterinary Manual (Kahn and Line 2005). Where information was incomplete, we conducted targeted literature searches on Scopus. To capture the full diversity of disease presentation, a separate record was created in the database whenever a source provided distinct information for a specific host taxon, causative pathogen, or host context (e.g., a specific WOAH entry for wildlife). For example, the single conceptual disease 'Toxoplasmosis' was expanded into four separate records to capture distinct clinical information documented for mammals and birds, in both wildlife and general/livestock contexts. This detailed extraction process resulted in a database of 295 unique records. The full dataset is available at https://doi.org/10.5281/zenodo.17467038.

3.2.2 Host and Disease Metadata

For each of the 295 records, we collected the following metadata.

Pathogen, transmission and zoonotic potential

The pathogen genus and species were recorded. To assess the relationship between visible signs and pathogen, each pathogen was assigned to one of seven taxonomic divisions: prion, virus, bacterium, protist, fungi, ectoparasite (parasitic arthropods), or helminth. We categorised the primary mode of transmission for each disease into one of three groups: direct (requiring close or direct contact between an infected and susceptible host), indirect (including transmission via fomites, ingestion, or environmental reservoirs), or vector-borne (requiring an intermediate organism, typically an arthropod, to transmit the pathogen). Zoonotic potential was sourced from the WOAH portal and supplemented by Taylor et al. (2001).

Host data

Hosts were classified into eight taxonomic groups: molluscs, insects, crustaceans, amphibians, reptiles, fish, birds, and mammals. To provide a more nuanced understanding of the host context, each disease-host association was further classified into one of three categories based on its primary source and relevance to wildlife. The first category, 'WOAH-designated Wildlife', comprises associations where the disease is formally designated by WOAH as being of significant importance to wildlife health and conservation. The second category, 'Other Wildlife Associated', includes associations not formally designated by WOAH for wildlife but for which a wildlife host was identified through supplementary literature sources. All other associations, primarily those concerning domestic animals, were classified as ‘Livestock / General’. This classification reflects the host contexts described in the sources reviewed; consequently, diseases in the 'Livestock / General' category may also affect wildlife.

Visible signs data

For each record, clinical signs were first extracted verbatim from the source texts. For example, the WOAH technical disease card for Elephant Endotheliotropic Herpesvirus states that typical clinical signs include "lethargy, anorexia (...) oedema of the head, neck, trunk, and thoracic limbs. Cyanosis of the tip of the tongue (...) and oral ulcers are also seen" (Bucko and Gieger 2019). Two researchers (Perkins and Olsen) then independently reviewed these textual descriptions to classify each of the 295 records as having either ‘visible’ or ‘non-visible’ signs. Initial independent agreement was observed for 96.6% of records

(n=285). The remaining ten disagreements all concerned diseases where diarrhoea was the only potentially visible sign; following discussion, we labelled these as ‘non-visible’. A record was classified as ‘visible’ if the text described at least one external, physical sign that would be observable on the animal itself. A record was classified as 'non-visible' if the signs described were purely behavioural (e.g., ‘lethargy’), were not affixed to the host's body (e.g., diarrhoea), or were secondary products (e.g., misshapen eggs). Thus, for the EEHV example, descriptions like ‘oedema of the head’, ‘cyanosis’ and ‘oral ulcers’ led to a ‘visible’ classification.

For further analysis, visible signs were grouped into eight categories based on their primary visual characteristic. These categories were: external exudate (e.g., nasal discharge, excessive mucus production); colour change (e.g., reddening of the body, dullness, gangrene); fur/skin/feather change (e.g., ruffled feathers, protruding scales, fin rot, fur loss); morphology (e.g., curling of the foot, protruding eyes); swelling/oedema (e.g., swollen joints, bloating, enlarged lymph glands); lesion (e.g., haemorrhaging, necrosis, ulcers, blisters); anorexia/weight loss (e.g., atrophy of the body); and conjunctivitis (e.g., swollen eyes, lachrymation, blepharitis). Applying this framework to the EEHV example, ‘oedema of the head’ was classified as swelling/oedema, ‘cyanosis of the tip of the tongue’ as colour change, ‘oral ulcers’ as lesion, and the associated ‘anorexia’ as anorexia/weight loss.

3.2.3 Assessing Image Availability

We used the ‘Photosearcher’ package (Fox et al. 2020) in R v4.2.2 (R Core Team 2021) to quantify images showing signs of diseases using open access photo sharing website Flickr. ‘Photosearcher’ relies on appropriate labelling so will only detect diseases that are labelled correctly with a text keyword. We use this method as a test bed to reflect the wider availability of images across the internet.

For each disease in our database, we compiled a list of relevant search terms, including the consolidated disease name based on the WOAH animal diseases database entry and the scientific names of any associated pathogens (see Supplementary Table 1 for a full list). This search did not separate different taxa or wildlife versus livestock. The total image count for each disease was calculated by summing the search results from all its associated terms. Our search included all images uploaded to Flickr between its launch (01/01/2004) and the

search date (22/01/2026), but with no spatial filters. To reduce inflation from ambiguous or widely used terms, we excluded a small set of search terms that returned large volumes of clearly irrelevant content on brief manual inspection (MERS, glanders, plague and mange).

3.2.4 Dataset Analysis

For statistical analysis, the 295 records from the Detailed Disease-Host Database were consolidated to our primary analytical unit: the unique disease-host association. This consolidation, which yielded a final dataset of 250 unique analytical units, was done to ensure that each data point in the model was statistically independent. When a disease-host pair had multiple entries (e.g., from different sources), we aggregated their metadata into a single summary record.

To investigate factors associated with the presence of visible signs, we fitted a binomial Generalised Linear Model (GLM). The binary response variable was the presence/absence (1/0) of visible signs. The model included the following predictors: host taxa (8-level factor), pathogen class (7-level factor), zoonotic (binary), wildlife disease (binary), WOAH-designated Wildlife (binary), and transmission routes. As a disease can have multiple transmission routes, we included direct, indirect, and vector transmission as separate binary predictors. We set 'mammals' and 'virus' as the reference levels for host taxa and pathogen class, respectively. We verified model assumptions using diagnostic plots from the 'performance' package (Lüdecke et al. 2021). To determine the overall significance of each predictor, we performed a sequential Analysis of Deviance (Type I) using a Chi-squared test. We then used the 'emmeans' package (Lenth 2024) to conduct *post-hoc* pairwise comparisons. All statistical analyses were done in R v4.3.2 (R Core Team 2021), using a significance level of 0.05.

3.3 Results

We found that 137 (67.2%) of the 204 infectious diseases listed in the World Organisation for Animal Health animal diseases portal have at least one visible sign. However, because many pathogens are generalists, a single disease can manifest differently in different hosts. Of the 250 disease-host associations, 162 (64.8%) had visible signs. Of these, 58 (35.8%) presented with a single category of sign, while the remaining 104 presented with two or more, with some showing as many as six distinct sign categories (Figure 3.2). After

consolidating all signs for each unique disease-host unit, the three most common categories of clinical signs were lesion (n=80), swelling/oedema (n=58), and colour change (n=58) (Figure 3.2).

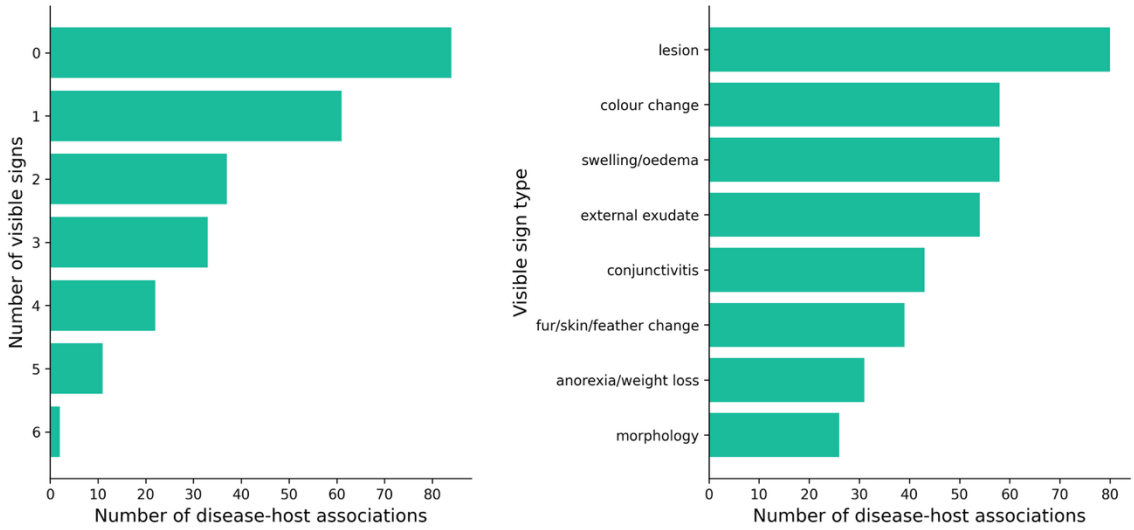


Figure 3.2. The frequency distribution of diseases with visible signs (left) and the type of visible signs (right) using data from the World Organisation for Animal Health animal diseases portal and additional sources.

Among the 162 disease-host associations with visible signs, viruses were the most common pathogen (n=86), followed by bacteria (n=37) and protists (n=19). Mammals were the most common host taxa (n=89), followed by birds (n=29) (Figure 3.3). Of these 162 disease-host associations, 79 (48.8%) involved a ‘WOAH listed’ disease. These visible associations were most commonly found in hosts categorised as 'Other Wildlife Associated' (n=94), followed by 'WOAH-designated Wildlife' (n=44), and 'Livestock / General' (n=24). Notably, of the 70 total associations involving 'WOAH-designated Wildlife', 44 (62.9%) had visible signs (Figure 3.3). Of the 137 unique diseases with visible signs, 55 (40.1%) are known to be zoonotic, compared to 31 (46.3%) of the 67 unique diseases without visible signs.

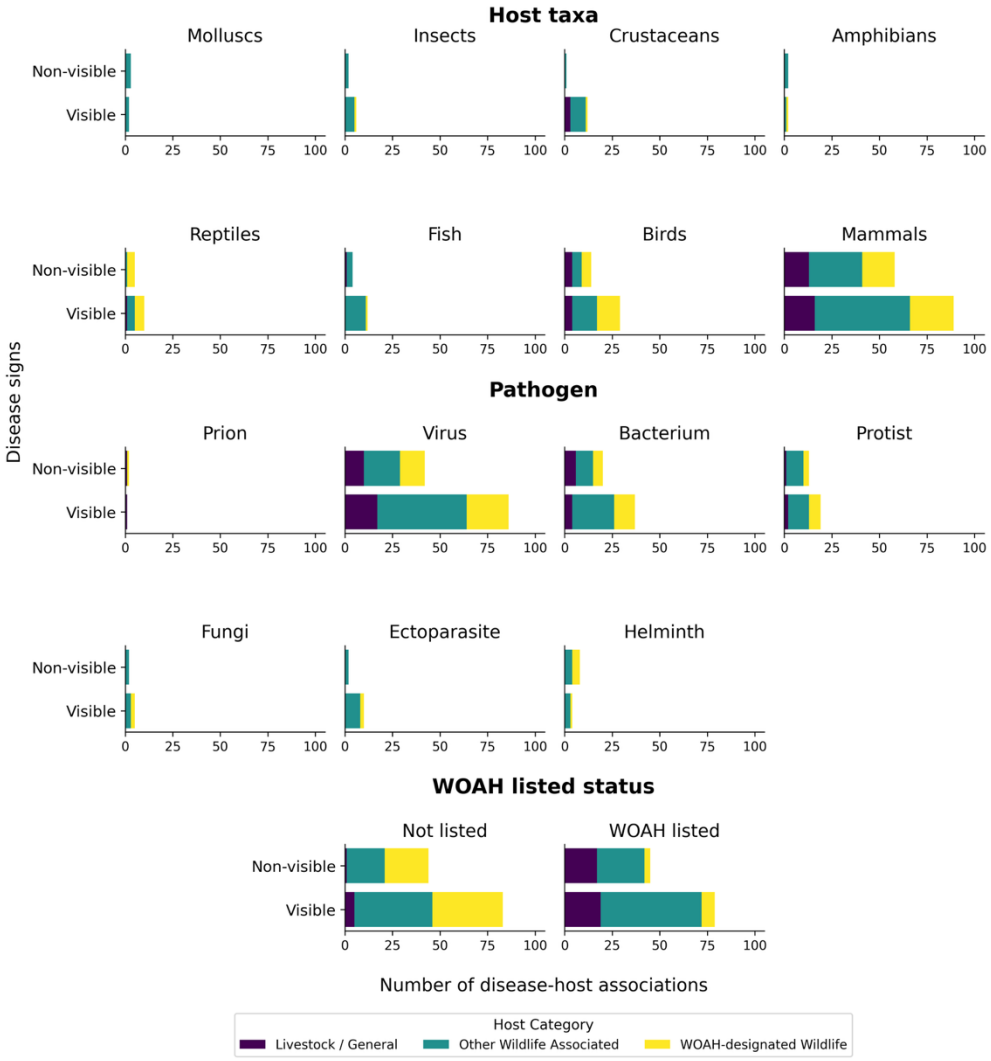


Figure 3.3. The number of diseases with visible signs, host category, and the number of diseases with non-visible signs according to host taxa (top) and pathogen class (middle), and World Organisation for Animal Health (WOAH) listed status (bottom) using data from the WOAH animal diseases portal and additional sources.

The majority of diseases in the WOAH database were viruses associated with mammals, of which over half had visible signs (Figure 3.4). Of all predictors tested, only direct transmission was significantly associated with the presence of visible signs ($\chi^2 = 7.72$, df = 1, p = 0.005). We found no significant association for host taxonomic group ($\chi^2 = 9.33$, df = 7, p = 0.230), pathogen class ($\chi^2 = 7.93$, df = 6, p = 0.243), zoonotic potential ($\chi^2 = 1.58$, df = 1, p =

0.209), or whether a disease affects wildlife (affects wildlife: $\chi^2 = 2.41$, df = 1, p = 0.120; WOAH_wild: $\chi^2 = 0.22$, df = 1, p = 0.638).

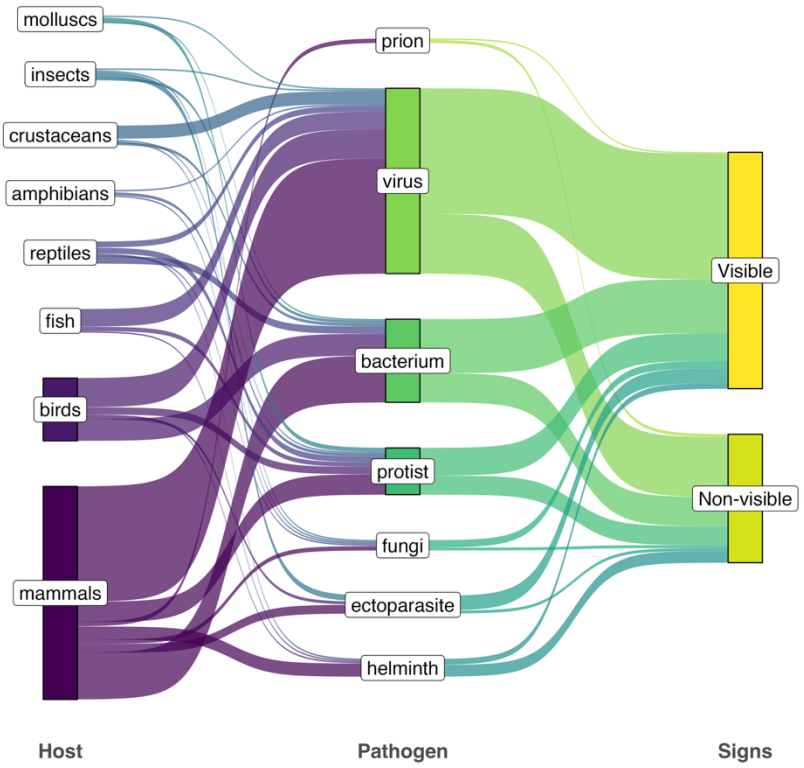


Figure 3.4. Co-occurrence and frequency of diseases associated with host, pathogen type and visible signs using data obtained from the World Organisation for Animal Health animal diseases portal.

These results confirmed that diseases with a direct transmission route were significantly more likely to present with visible signs (Estimate = 0.94, p = 0.004). No other individual predictor level was statistically significant, although diseases in crustaceans showed a marginal, non-significant trend towards being more likely to have visible signs compared to the reference taxon, mammals (Estimate = 2.05, p = 0.063). Full model coefficients are provided in Supplementary Table 2. Post-hoc pairwise comparisons for both host taxa (Supplementary

Table 3; Supplementary Figure 2) and pathogen class (Supplementary Table 4; Supplementary Figure 3) confirmed that no single pair of levels was significantly different from another.

3.3.1 Assessing Image Availability

To quantify online image availability, we searched Flickr for the 137 diseases identified as presenting with visible signs. We found images for 74 of them, although the number of images found for each disease varied greatly. Six diseases returned over 1,000 images, 20 had between 100 and 999 images, 18 had between 10 and 99, and the remaining 30 had fewer than 10 images (see Figure 3.5, Supplementary Table 5). Because Flickr counts are keyword-based and some terms were dominated by non-disease content on brief manual inspection, we excluded a small set of ambiguous terms (MERS, glanders, plague and mange) from the summary; in the filtered results, diseases with the most images included malaria (n=5,018), anthrax (n=3,789) and pox virus infections (n=3,147).

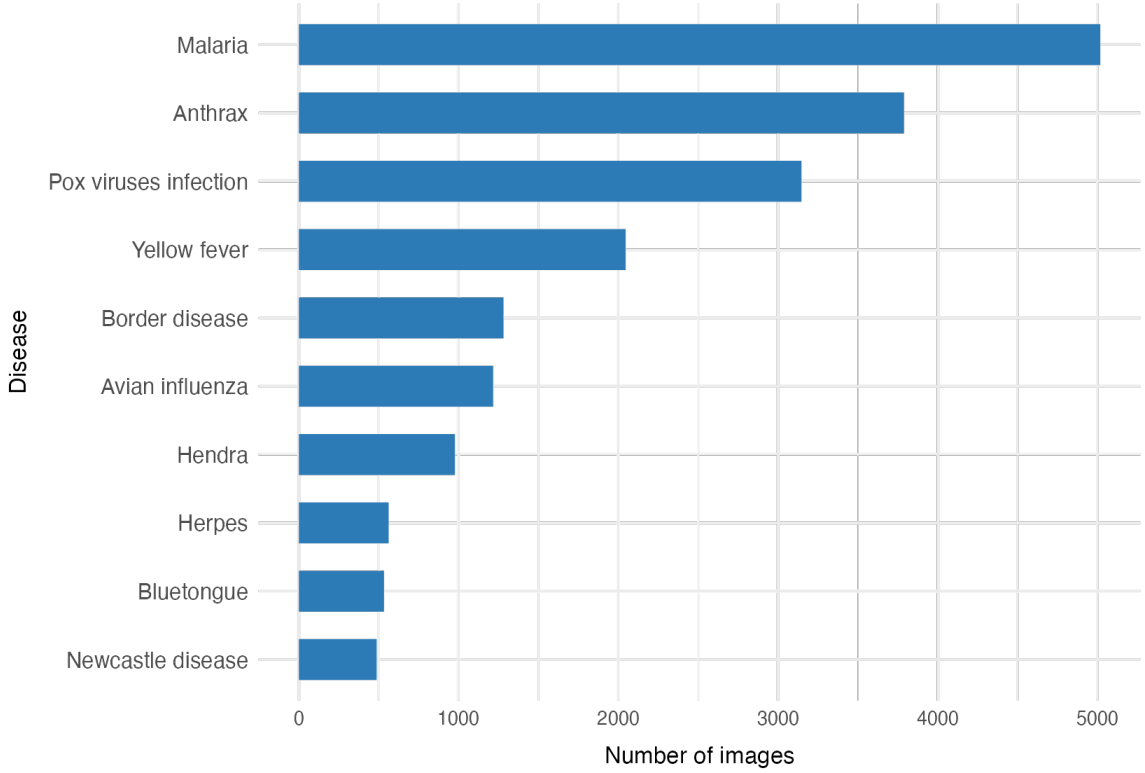


Figure 3.5. The top 10 diseases with the highest image availability on Flickr, out of the 137 diseases from the World Organisation for Animal Health animal diseases portal identified as

having visible signs. The number of images for each disease is based on a search conducted on 22 January 2026.

3.4 Discussion

Our study shows that a substantial proportion (64.8%) of the 250 disease-host associations of animals that we evaluated have visible signs of infection, confirming a broad potential for image-based surveillance. These signs were most frequently found in diseases of mammals caused by viruses, and the most common signs were lesions, colour changes, and swelling. This initial classification provides a foundation for identifying suitable diseases and visual characteristics that could be targeted by automated detection models.

Our statistical model revealed that after accounting for host and pathogen type, the only significant predictor of sign visibility was transmission mode. This link between direct transmission and visible signs aligns with the evolutionary framework of "transmission-symptom" co-evolution (Antonovics et al. 2017). The theory posits that for pathogens reliant on direct contact, external signs are not just by-products of infection but can evolve to be integral to the transmission process itself. This may occur because signs that facilitate shedding, such as lesions, increase the probability of a pathogen transferring to a new host. Our data support this: two of the most frequent signs in directly transmitted diseases were lesions and external exudates. Such signs can create a direct physical pathway for a pathogen to exit an infected host and enter a susceptible one, for example via infectious fluid from a sore. In this context, the visible sign could be a key component of the pathogen's transmission route.

Our model found no significant link between visible signs and pathogen class, host taxa, or zoonotic potential. These results, however, should be interpreted cautiously. Because we used the disease-host pair as our analytical unit, some pathogen groups, particularly generalist viruses, were over-represented in the dataset, which may have obscured weaker underlying associations.

While images offer opportunities for disease surveillance, the success of automated detection depends not only on how conspicuous a sign is, but also on the diversity of its presentation. Signs vary in appearance across hosts, individuals, stages of infection, and environmental conditions. Our model showed no significant overall effect of host taxa but, for some, the sample size was probably too small e.g. just 13 crustacean specific diseases. In early-stage infections, visible signs such as slight fur thinning or faint skin discolouration may be too subtle to detect reliably. Signs like fur loss from mange are often stark and manual image-based disease detection has worked well (Scott et al. 2020; Murray et al. 2021). Other signs, however, like minor texture changes or slight swelling, can be subtle and easily missed. Although 'external exudate' was a common sign, and might seem visually subtle, it has been used to detect Foot and Mouth disease from images of 'ocular discharge' (Hofstra et al. 2023). Regardless, image-based surveillance will likely perform best for diseases with bold, unmistakable visual hallmarks.

Most diseases in our study are not pathognomonic, so an image with visible signs cannot provide a definitive diagnosis. They can, however, be used for disease surveillance, defined by WOAH and IUCN (2024) as observing clinical signs to detect potential disease events. Using images as non-biological samples is a non-invasive method for health assessment, distinct from pathogen surveillance which requires collecting biological samples from captured animals. Image-based methods may be less sensitive for detecting subtle, early-stage infections, but they offer a low-cost, scalable way to screen for disease where no surveillance previously existed. This approach can also be used to assess historic disease emergence or outbreaks. For example, Elmer et al. (2019) used archival images from Google to trace the historical occurrence of black spot syndrome in Caribbean reef fishes back to 1985, demonstrating the value of public image repositories for long-term disease surveillance. The methods for acquiring these non-biological samples align with established surveillance strategies. Using public websites like Flickr or iNaturalist is a form of passive and participatory surveillance, which relies on opportunistic reports from citizen scientists to detect mortality events and unusual signs (Lawson et al. 2015). In contrast, using camera-trap networks (e.g. Ringwaldt et al. 2025) or dedicated video transects (e.g. de Wit and Johnson 2024) to monitor specific diseases is a form of active (targeted) surveillance.

A key part of image-based surveillance is acquiring the images. In a livestock setting, Precision Livestock Farming is growing the capability to collect many images for real-time health monitoring (Berckmans 2017). The commercial aquaculture and fisheries sectors also use automated imaging to monitor sea lice and track wild fish stocks (Skirrow 2024; Tidal 2025). Acquiring quality images in wildlife settings provides other challenges, but aligns with a broader trend towards non-invasive methods in wildlife health research (Schilling et al. 2022). Here, we evaluated Flickr as a proxy for open-access imagery and found data for over half of the diseases, affirming the potential of public photo-sharing platforms for filling in gaps in disease surveillance. Our search returned substantial image counts for several diseases with more specific terminology (e.g., anthrax, avian influenza, foot and mouth disease, and chytridiomycosis; Supplementary Table 5), suggesting that for some conditions sufficient public imagery may already exist to support model development. However, term choice matters: broad or ambiguous search terms can return large volumes of mixed-content imagery. For example, the term “mange” returned 42,260 images, but a brief manual check suggested that many results were off-target, implying that such datasets would require substantial verification and curation. A practical way forward is to use more specific, context-rich queries (e.g., combining host and disease terms such as “fox” + “mange”), which yields a more manageable candidate set. However, even with improved queries, verification remains a standard requirement in wildlife disease research. For example, Scott et al. (2020) manually analysed images sent in by the public to study mange, while Ringwaldt et al. (2025) used computer vision on dedicated camera trap images. Similarly, Szentivanyi and Vincze (2022) manually screened images from iNaturalist to track toad myiasis.

Future image-based disease surveillance requires integrating these strategies within a One Health framework. Further work should focus on creating better analytical pipelines for public datasets and encouraging more targeted data collection. For example, stakeholder groups like recreational anglers could help build expertly labelled training sets for fish diseases. Existing ecological camera-trap networks could also be repurposed for health monitoring. As accessible computer vision pipelines emerge (e.g. Blair et al. 2024; Brook et al. 2025) these methods become more feasible for a wider range of users. Combining image

data with other metadata, such as geolocation and environmental variables, can generate data to better inform One Health actions.

3.5 Conclusion

Our findings confirm that a majority of animal diseases in the WOAH animal diseases database present with externally visible signs, establishing a broad potential for image-based surveillance. Realising this potential requires strategies that navigate the complexities of disease ecology and the inherent noise in public image datasets. Our work aligns with the principle that diseases with visible signs are ideal targets for citizen science (Lawson et al. 2015) and a wider trend towards non-invasive sampling in wildlife research (Schilling et al. 2022). While not a replacement for traditional diagnostics, these methods provide a valuable tool for scalable, low-cost screening. For many diseases in remote settings where surveillance is sparse, this approach can provide a mechanism for detection where previously there was none. By combining proactive data collection with advanced computer vision and integrating these efforts into existing One Health frameworks, image-driven tools can improve near-real-time disease monitoring and contribute to more timely interventions.

CHAPTER 4 :

THE EXTENDED IMAGE: THE VALUE OF ONLINE IMAGES FOR ECOLOGICAL RESEARCH

Authors: Agnethe Olsen[1], Chris B. Jones[2], Paul Rosin[2], Neil Cook[1], Jo Cable[1], Sarah E. Perkins[1]

Affiliations:

[1]School of Biosciences and Water Research Institute, Cardiff University, CF10 3AX, UK.

[2]School of Computer Science and Informatics, Cardiff University, CF24 4AG, UK.

Abstract

Online digital images are a large and growing data source that can supplement traditional ecological monitoring. Here we extend the concept of the 'extended specimen', traditionally applied to physical specimens with associated data network, to include digital images. Photographs reflect species occurrence, but we propose further ecological utility, such as health indicators. Here, we use fish as a model species to assess image availability and 'extendibility'. Fish are globally important with economic and cultural value, and many images are shared online by anglers, making them a good case study. We systematically assessed quantity, quality, and accessibility of ~70,000 salmonid images from four major online repositories (iNaturalist, Flickr, GBIF, Wikimedia Commons). iNaturalist contributed the most verified salmonid images with robust spatial and temporal metadata, followed by GBIF for metadata completeness. Species-level identification was common, but reliability varied; iNaturalist and GBIF offer structured verification, whereas taxonomy in Flickr and Wikimedia Commons was derived from image search terms. We found strong support for health monitoring, with visible signs of disease or damage in 17.4% of an annotated subset of images. However, we identified key limitations for 'extended' analyses; restrictive licensing hindered data reuse, and a lack of standardised size references made estimating fish size impractical. This demonstrates the potential for opportunistic surveillance and that repositories require evaluation of characteristics including data quality, licensing, and taxonomic verification. Nevertheless, leveraging the distinct strengths of diverse platforms, including using images with incomplete metadata for AI model training, is critical to maximise 'extended image' ecological insights.

4.1 Introduction

4.1.1 Digital Images as an 'Extended Specimen'

The current global biodiversity crisis has severe implications for environmental and societal wellbeing. Understanding species distributions is critical for biodiversity and environmental management to mitigate these impacts (Pecl et al. 2017; Pyšek et al. 2020). Traditional monitoring is, however, resource-constrained (Yoccoz et al. 2001; Lindenmayer and Likens 2010; Amano et al. 2016), with systematic data often undermined by spatial, temporal, and taxonomic gaps (Isaac and Pocock 2015; Speed et al. 2018). Here, we propose the 'extended image' framework, harnessing massive image databases as a step-change to address this.

Digital technology has greatly increased ecological data availability, with citizen science in particular collating image-based records at unprecedented scales (Nazir and Kaleem 2021; Depauw et al. 2022). Dedicated platforms like eBird (eBird.org) and iNaturalist (iNaturalist.org) have accumulated vast numbers of biodiversity records, many with photographic evidence. In addition to these structured programmes, a large repository of images exists on other online platforms, uploaded without specific scientific intent. The use of such digital sources, generated for other purposes, to quantify ecological patterns and processes is described as iEcology (Jarić et al. 2020). Within this broader iEcology space, 'passive citizen science' (see Edwards et al. 2021) specifically leverages social media content that is unconnected to any particular citizen science programme, treating everyday uploads as an untapped dataset for studying species trends, distributions, behaviour, and/or phenology (Vardi et al. 2024). This parallels with the natural history concept of the 'extended specimen', where physical specimens are anchors for a rich network of associated digital data (Webster 2017; Lendemer et al. 2020). Building on this, the 'Digital Extended Specimen' (DES) framework aims to formalise these as interconnected, FAIR (Findable, Accessible, Interoperable, and Reusable) digital objects on the internet (Hardisty et al. 2022).

Analogously, we can conceptualise the 'extended image' where all associated data are linked to the image itself, rather than to a physical specimen. Online photographs and videos are not merely species occurrence records but potential repositories of diverse ancillary data such as behaviour, interactions, or health status. For example, species distribution data from Flickr closely mirrors National Biodiversity Network Atlas records (Edwards et al. 2021), and online photos enabled the rediscovery of a snake species

previously thought extinct (Durso et al. 2021). Plant-pollinator interactions have been extrapolated from citizen science wasp imagery (Pernat et al. 2024), YouTube videos opportunistically revealed rare thanatological behaviours in Asian elephants (*Elephas maximus*) (Pokharel et al. 2022) and online images have been used to document terrestrial hermit crabs using artificial shells (Jagiello et al. 2024). The 'extended image' is also relevant for wildlife health surveillance; the severity and drivers of Black Spot Syndrome in ocean surgeonfish (*Acanthurus tractus*) have been investigated from videos (de Wit and Johnson 2024), while photographs from iNaturalist and Flickr can track nasal toad myiasis (Szentivanyi and Vincze 2022). Applying these approaches to fish disease surveillance is further explored in our own work (Olsen et al. submitted; Chapter 5).

4.1.2 Applying the 'Extended Image': a Salmonid Case Study

Fish are an ideal taxonomic group to explore extended image potential: global recreational anglers number hundreds of millions (Arlinghaus et al. 2020), and many using digital tools to share catch information generating a substantial repository of images with associated ecological and social insights (Monkman et al. 2018; Skov et al. 2021; Lennox et al. 2022). Salmonids warrant particular research attention, as keystone species and bioindicators (Naiman et al. 2002; Schindler et al. 2003) with substantial economic value and deep cultural importance, yet increasingly under environmental pressures (Reid et al. 2019). The value of the 'extended image' is pertinent for fish health, as externally visible disease signs are common (Noga 2010) and could be captured in publicly available images. Indeed, computer vision (CV) for monitoring salmon populations is being explored (e.g. Atlas et al. 2023; Olsen et al. in review). Since infectious diseases threaten both wild and farmed salmonid populations (Peeler et al. 2011; Thorstad et al. 2021), disease monitoring through scalable image analysis could provide critical policy and risk mitigation insights (Olsen et al. in review). Understanding fish distribution and health, therefore, has conservation, socio-economic, and food security relevance (Lynch et al. 2016; FAO 2020).

The value of online images for biodiversity monitoring is recognised (Lynch et al. 2016; FAO 2020), yet no systematic assessment has evaluated availability, quality, or accessibility of fish images across multiple platforms for comprehensive ecological analysis, i.e. 'extendibility'. We therefore aimed to address this knowledge gap using salmonid images

from four major open-access platforms. We conducted a systematic assessment of temporal patterns, geographical coverage, taxonomic reliability, licensing restrictions, and the prevalence of visible signs of disease or damage. Examining both structured citizen science platforms and informal image-sharing websites, we aimed to identify strengths and limitations of different data sources to guide researchers using online images as an ecological tool. We hypothesised that online platforms would yield substantial and geographically diverse salmonid image datasets. We further predicted that these would be suitable for addressing ecological questions regarding distribution, seasonal and annual occurrence patterns, and health.

4.2 Methods

4.2.1 Data Source Descriptions

Wildlife images are available from many sources, including social media feeds, angler fora, and institutional archives. For this study, however, we focused on four major open-access platforms designed for broad image sharing and/or biodiversity data aggregation: iNaturalist (inaturalist.org), GBIF (gbif.org), Wikimedia Commons (commons.wikimedia.org), and Flickr (flickr.com). These were selected for their high expected volume of relevant images and the availability of metadata, which is often removed on general social media platforms (Steidl 2016). These metadata are essential for developing the 'extended image' for ecological analysis.

iNaturalist is a citizen science platform with over 260 million observations from approximately 9 million users as of July 2025. GBIF (Global Biodiversity Information Facility) aggregates research data, with over 3.1 billion occurrences and 239 million images. Wikimedia Commons contains over 123 million freely usable media files. Flickr is a photo-sharing website hosting an estimated 10 billion images and is popular for wildlife photography. Since 2019, Flickr has limited free accounts to 1,000 photos, which may affect the availability of older content (Gartenberg 2018). For platforms without taxonomic organisation (Flickr and Wikimedia), scientific and common names was compiled by integrating data from FishBase, FishTreeOfLife, and NCBI. Common names in English were sourced from NCBI, FishBase, and iNaturalist (Supplementary Table 6).

We also originally considered other image sources. Previous research have used images from platforms like X (formerly Twitter) and Facebook (Daume 2016; Durso et al. 2021) but ultimately these were excluded due to API restrictions and subscription fees, which create significant financial and technical barriers. In addition, these platforms remove EXIF data from uploaded images restricting their value in ecology. CalPhotos, with 188,152 animal images, was also excluded due to the absence of an API and requirements for individual image download permissions.

4.2.2 Data Acquisition, Verification, and Initial Dataset Compilation

We collected images and metadata (e.g., date, location, user-provided tags and descriptions, species identification, image licensing information, EXIF data) for the Family Salmonidae from the selected platforms between December 2023 and February 2024. All data acquisition and processing scripts were run using Python (v3.11.7) unless otherwise specified.

For iNaturalist, we downloaded metadata files from the ‘iNaturalist Licensed Observation Images’ dataset on 4 December 2023 using the AWS Command Line Interface. Using pandas (v1.5.3) in Python, we filtered for observations classified as Salmonidae (taxon_id = 47520) or lower taxonomic levels, which returned 49,040 unique image URLs. Of these, 40,569 were "Research Grade", indicating community consensus on the species identification for an observation with date and coordinates. These images were downloaded on 12 December 2023 using the multiprocess package (v0.70.15); six images were unavailable due to removal from the platform since the metadata download.

From GBIF, we searched using the following filters: BasisOfRecord is Human Observation", "MediaType is Image", "OccurrenceStatus is Present", and "TaxonKey is Salmonidae".
To avoid duplication, we excluded the 94.5% of records that originated from iNaturalist. This filtering resulted in a Darwin Core Archive containing 1,843 unique images from other sources such as museum collections, which were subsequently downloaded using the gbif-dl package (v0.1.1) in Python (v3.9.18).

Using an adapted R script from Marshall et al. (2020) , we queried our compiled list of salmonid names (Supplementary Table 6) to retrieve 183 unique images from Wikimedia Commons via Python's urllib3 (v1.26.16).

For Flickr, we queried the API on 12 December 2023 using the flickrapi package (v2.4.0) in Python. As the platform lacks formal taxonomic identification, we searched image titles, descriptions, and user-supplied tags using a curated list of target names (see Supplementary Table 6).

The search query was refined through preliminary testing to remove unproductive terms. For example, the term 'Peled' was excluded as it returned images of the footballer Pelé but no relevant images of the fish, *Coregonus peled*. The final, refined list of 43 search terms yielded 19,610 licensed images. To remove exact digital duplicates arising from multiple user uploads or tags, we generated a unique SHA-256 hash for each image file using Python's Hashlib library. This deduplication resulted in 18,101 unique images, although this method does not account for near-duplicates such as sequential photographs from a single event. When a unique image had multiple metadata records, we consolidated them by retaining the record with the highest location accuracy and the earliest available date.

The combined dataset comprised 69,167 unique images. However, the images from Flickr and Wikimedia Commons still required manual verification. This was necessary to filter results from ambiguous search terms that, unlike 'Peled', did yield a mixture of relevant and irrelevant images. For example, a search for 'Grayling' returned images of the target fish (Thymallus spp.) but also the non-target Grayling butterfly (*Hipparchia semele*). An initial sift of the 18,284 images from Flickr and Wikimedia was performed by two annotators to remove images that were obviously not fish. Following this, 9,773 images were evaluated in detail on Labelbox by at least two annotators, who assessed for the presence of a fish and confirmed it as a salmonid. We excluded images of cartoons, drawings, prepared food, or fossils (see Figure 4.1 for examples of included and excluded images). This process verified 91 of 183 images from Wikimedia Commons and 6,869 of 18,101 images from Flickr as suitable for the final dataset.

Figure 4.1. Examples of image search results highlighting classification challenges. (A) Atlantic salmon (Salmo salar). (B) T-shirt with salmon design. (C) Arctic grayling (*Thymallus arcticus*). (D) False positive: Grayling, a butterfly species, correctly labelled but incorrectly returned in search for salmonids due to taxonomic name overlap. Image credits: (A) "Atlantic Salmon", © E. Peter Steenstra/USFWS, Public Domain; (B) "Student-made Atlantic salmon T-shirts", © Gillian Ball / USFWS, Public Domain; (C) "The next generation's parents?", © USFWS Mountain-Prairie, some rights reserved (CC BY); "Tree Grayling. *Neohypparchia statilinus*", © gailhampshire, some rights reserved (CC BY)

4.2.3 Metadata consolidation

To enable robust cross-platform analysis, we standardised key metadata fields across all sources, as detailed in Supplementary Table 7. For iNaturalist and GBIF, taxonomic classification was taken directly from the metadata fields provided in the downloaded datasets. For Wikimedia Commons and Flickr, taxonomic information was assigned based on the search term used to download the image; if an image was downloaded under multiple terms, the lowest available taxonomic level was assigned. Where images were downloaded using a common name (e.g., "European grayling"), the corresponding scientific name and taxonomic level were determined using Supplementary Table 6.

Quality indicators were harmonised under a standardised 'quality' column: iNaturalist provided structured quality grades, GBIF included an ‘identificationVerificationStatus’ field, and Flickr and Wikimedia Commons relied on manual verification labels. Spatial information was standardised under 'longitude', 'latitude', and 'location_accuracy'. Flickr provided coordinate data with an accuracy scale, while iNaturalist and GBIF expressed uncertainty as a radius in metres. Wikimedia Commons generally lacked spatial metadata. Temporal data were harmonised under a single 'date' field, drawing from observation, upload, or EXIF dates.

Licensing information was mapped to ten standard categories to ensure consistent tracking of usage rights. These included All Rights Reserved, Public Domain, Government Work, Unknown, and six Creative Commons (CC) variants (CC BY, CC-BY-SA, CC BY-ND, CC BY-NC, CC BY-NC-SA and CC BY-NC-ND). The CC licenses specify different permissions, such as restricting for-profit use (-NC), preventing modifications (-ND), or requiring derivative works to be shared under the same license (-SA). Data provenance was maintained by preserving each image’s unique source identifier and repository of origin. In addition to these standardised fields, GBIF provided the richest ecological metadata through the Darwin Core standard. This included fields such as 'habitat', 'samplingProtocol', 'lifeStage', 'behaviour', 'establishmentMeans', and 'organismQuantity'. Further contextual information, such as field notes, taxonomic details, and environmental parameters (e.g., 'minimumDistanceAboveSurfaceInMeters'), was also available for some records. These additional fields were retained for reference and potential future analyses.

4.2.4 Dataset Characterisation: Image Quality and Licensing

Image Quality Assessment

Image Quality Assessment (IQA) models are computational tools designed to predict the perceived quality of an image. No-reference (or "blind") IQA models operate solely on the input image, whereas full-reference models require a comparison to a pristine reference image (Ding et al. 2021). To evaluate the quality of images collected for this study, we applied three established no-reference methods: blur detection (Pech-Pacheco et al. 2000), BRISQUE (Mittal et al. 2012), and NIQE (Mittal et al. 2013). We selected these methods to provide a comprehensive characterisation of the dataset. Blur detection

specifically measures image sharpness, which is critical for the visual identification of diagnostic features. BRISQUE and NIQE assess overall perceptual quality and are sensitive to a wide spectrum of distortions, including noise, compression artefacts, and other deviations from natural scene statistics.

Blur scores were calculated using the Laplacian variance method, which quantifies sharpness by computing the variance of the image's Laplacian operator. BRISQUE scores were obtained using the QualityBRISQUE class from the OpenCV library (opencv-contrib-python 4.9.0.80). NIQE scores were calculated using a publicly available Python implementation which also relies on OpenCV. Notably, this implementation of NIQE requires a minimum image resolution of 192 × 192 pixels due to its patch-based analysis.

Because IQA metrics were non-normally distributed and sample sizes were highly unbalanced among platforms, we compared metric distributions among repositories using Kruskal–Wallis rank-sum tests and reported epsilon-squared () as an effect size. Where global tests were significant, we used pairwise Mann–Whitney U tests with Holm correction for multiple comparisons and reported rank-biserial correlation as an effect size. NIQE scores could not be calculated for a small number of images (<0.3%) that did not meet minimum requirements (e.g. resolution constraints), and these were excluded from NIQE-specific tests.

License Information Analysis

To assess the reusability of the compiled salmonid dataset, we analysed the distribution of image licenses. This analysis is essential as license agreements dictate how images can be used, reproduced, modified, and shared for research, including in publications and for training computational models. Our analysis categorised licenses based on their restrictions; for example, ‘All Rights Reserved’ typically prohibits reuse without explicit permission from the copyright holder. We analysed the license distribution across all source platforms for verified salmonid images. This approach was designed to quantify the dataset’s potential for scientific sharing and reuse, aligning with best practices for data stewardship and addressing known challenges in citizen science data openness (Groom et al. 2017).

4.2.5 Analysis of Spatial and Temporal Patterns

Location Information Analysis

The spatial distribution of verified salmonid observations was visualised using hexagonal binning. The visualisation included all images of salmonids with coordinate data, which were drawn from all sources except Wikimedia Commons as it generally lacks location-specific metadata. Location uncertainty was recorded differently across platforms. Flickr uses a 16-point accuracy scale ranging from 1 (world level) to 16 (street level), with intermediate values representing broader regions (e.g., country at level ~3, city at level ~11). Notably, Flickr defaults to the highest accuracy (level 16) when a user does not specify a value. In contrast, iNaturalist and GBIF express uncertainty as a radius in metres, defining the circular area in which the observation occurred.

Datetime Information Analysis

The temporal distribution of salmonid observations was analysed across the four image repositories. To ensure data quality, observations were filtered to event dates between 1 January 1900 and 14 December 2023, which removed five records (<0.01%) with apparent data entry errors. Following manual inspection, a small cluster of Flickr images dated to 1980 and attributed to two specific users were also identified as clear outliers and removed. We assessed annual trends by tracking the number of observations per year and evaluated seasonal patterns by aggregating observations by month across all years. To statistically support seasonal descriptions, we tested (i) whether monthly observations deviated from a uniform distribution using a Chi-square goodness-of-fit test (overall and per platform), and (ii) whether month-by-platform patterns differed among repositories using a Chi-square test of independence; we reported Cramér's V as an effect size and inspected standardised residuals to identify months contributing most strongly to differences.

4.2.6 Analysis of taxonomic composition

We analysed the taxonomic composition of the verified salmonid observations using Python. The taxonomic resolution was assessed across eight hierarchical levels (family, subfamily, genus, species, hybrid, subspecies, variety, and form), and the relative proportion of observations at each level was calculated for each data source. For a more detailed analysis, observations were organised into the three salmonid subfamilies (Coregoninae,

Salmoninae, and Thymallinae), with each genus assigned to its subfamily based on established taxonomic classifications.

4.2.7 Image Content Annotation: Size and Health Indicators

We annotated a subset of 5,667 verified salmonid images (11.5% of the total) to check for signs of disease or damage and for objects to estimate size. We selected this subset using stratified random sampling by data source and the lowest available taxonomic classification of genus or species to represent the different platforms and taxa. This mixed-level approach ensured the inclusion of images where species-level identification was not possible, which is a frequent occurrence in images of diseased or damaged fish. For this assessment, we sampled all verified salmonid images from Flickr, Wikimedia and GBIF. We sampled all iNaturalist images regardless of their quality grade. We did this because fish with visible disease or damage can be harder to identify to species level, and excluding images that were not 'Research Grade' could create a selection bias. Two annotators independently examined the image subset in Labelbox. To check for disease or damage, they recorded visible signs such as rashes, discolouration, parasites, growths, cuts, or wounds. We recorded an image as positive if any fish within it showed these signs. To find objects for size estimation, the annotators recorded standard references like rulers or coins but ignored non-standard objects like hands or fishing rods. We excluded images from the final analysis if poor quality or obstruction meant we could not reliably assess them.

4.3 Results

4.3.1 Dataset Overview: Image Quantity, Quality, and Licensing

Our cross-platform search and filtering process yielded a final dataset of 49,372 verified salmonid images (Table 4.1). This dataset was distilled from an initial collection of 69,167 unique, deduplicated images. The verification pipeline was essential, removing 10,711 entirely irrelevant (non-fish) images and a further 9,084 images of other fish species. This highlights a key challenge of using platforms without strict taxonomic controls. The composition and metadata of the final dataset varied significantly between platforms. iNaturalist was the primary contributor, providing the vast majority of verified salmonid images (40,569) with high-quality taxonomic data. In contrast, Flickr provided the highest

resolution images (median width 2560px) but had poor coverage for location data. GBIF offered the most complete metadata, with 100% of its images having date, time, and location information, though its images were the lowest resolution (median width 1000px).

Table 4.1. Summary of image counts and metadata availability across repositories. The table shows the number of images at each stage of the data collection and filtering pipeline. Columns illustrate the progression from initial downloads ('Total images'), through deduplication ('Unique images'), to manual verification ('Fish images' and 'Salmonid images'). * The 'Fish images' and 'Salmonid images' counts are subsets of the 'Unique images' total. ** All subsequent metadata and dimension statistics are calculated from the final 'Salmonid images' set.

Source	**Total images**	**Unique images**	**Fish images***	**Salmonid images***	**With Datetime****	**With location****	**Median image width [px] (IQR)****	**Median image height [px] (IQR)****
iNaturalist	49,057	49,040	49,040	40,569	40,056	40,568	1861 (1536-2048)	1536 (1365-2048)
Flickr	19,610	18,101	7,479	6,869	6,865	1,211	2560 (1181-3870)	1944 (890-2848)
GBIF	1,843	1,843	1,843	1,843	1,843	1,843	1000 (800-1280)	752 (600-1054)
Wikimedia Commons	183	183	94	91	54	0	1600 (800-2303)	965 (538-1732)
Total	70,693	69,167	58,456	49,372	48,818	43,622	-	-

The quality of salmonid images differed among repositories when assessed using three no-reference IQA metrics (Figure 4.2). Median blur scores (Laplacian variance; higher values indicate sharper images) ranged from 199.35 (iNaturalist) to 322.35 (GBIF), with Flickr (216.20) and Wikimedia Commons (231.57) intermediate (Figure 4.2a). Median BRISQUE scores suggested slightly better perceptual quality in GBIF (20.84) than in

Wikimedia Commons (22.08), Flickr (26.10), and iNaturalist (26.81), although all medians were <30 (Figure 4.2b). Median NIQE scores were tightly clustered (12.22–13.89), with Wikimedia Commons lowest (12.22), GBIF 12.92, Flickr 13.33, and iNaturalist 13.89 (Figure 4.2c). Despite substantial within-platform variation (Figure 4.2), non-parametric tests detected differences among platforms for all three metrics (Kruskal–Wallis, p<0.001), but the magnitude of these differences was small (=0.0038–0.0071). Post-hoc pairwise comparisons (Holm-adjusted Mann–Whitney tests) showed that sharpness differed among iNaturalist, Flickr, and GBIF (all adjusted p<0.001), whereas Wikimedia did not differ from other platforms; BRISQUE differed among most platform pairs except GBIF versus Wikimedia; and NIQE differed only for iNaturalist versus the other platforms, with no detectable differences among GBIF, Flickr, and Wikimedia after correction.

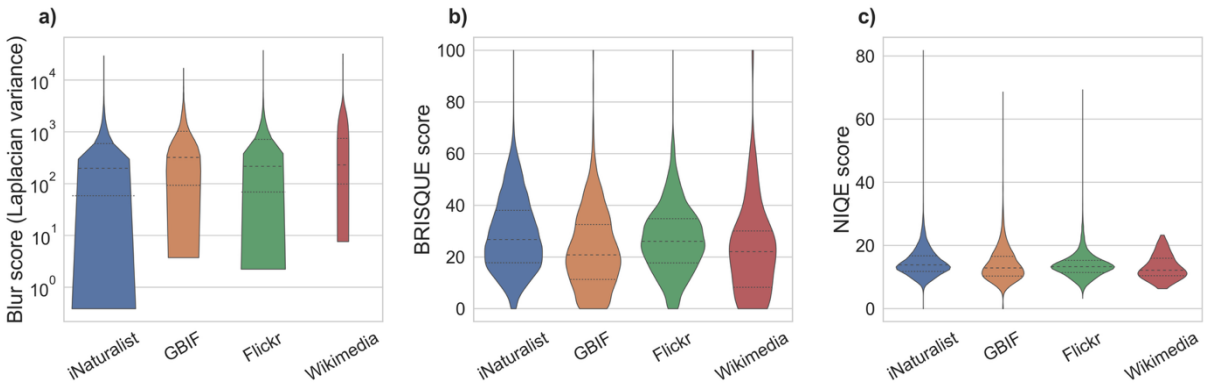


Figure 4.2. Variation in image quality metrics across repositories for verified salmonid images (n=49,372). Violin plots show the distribution of (a) blur score (Laplacian variance, a sharpness proxy; higher values indicate sharper images, shown on a log scale), (b) BRISQUE, and (c) NIQE (lower values indicate higher perceptual quality for BRISQUE/NIQE). Internal horizontal lines indicate the median and interquartile range. NIQE scores could not be computed for a small subset (<0.3%).

Two license types dominated the dataset of confirmed salmonid images. Most images (73.4%; 36,252) had a Creative Commons Attribution-NonCommercial (CC BY−NC) license. The 'All Rights Reserved' license was the second most common (11.6%; 5,736 images). The

proportion of license types differed between platforms (Figure 4.2). Most iNaturalist images were CC BY−NC (86.3%), while most Flickr images were 'All Rights Reserved' (83.5%). Wikimedia Commons had a high proportion of Public Domain and Creative Commons Attribution-ShareAlike (CC BY−SA) licenses, which together accounted for 81.4% of its images. GBIF images were mainly CC BY−NC (56.4%) and Creative Commons Attribution (CC BY) (25.4%).

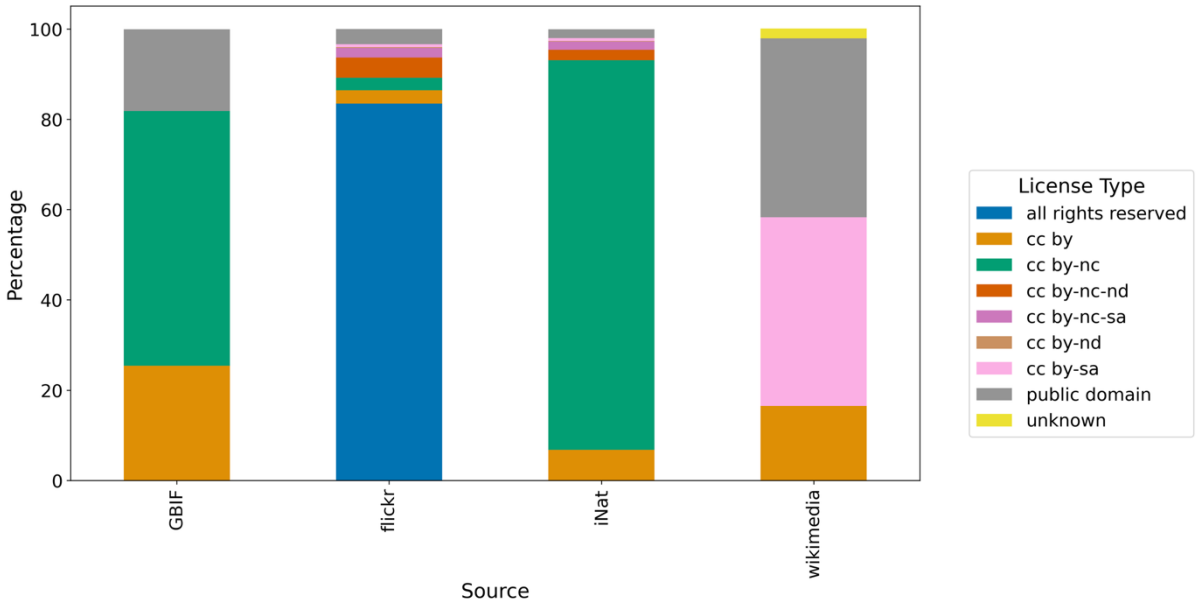


Figure 4.2 Image license distribution across data sources for confirmed salmonid images (n=49,372), shown as percentages per source. Stacked bars represent proportion of different license types within a source platform.

4.3.2 Spatial and Temporal Distribution

The precision of location data for the salmonid images varied between the sources that provided coordinates (Supplementary Figure 5). For iNaturalist and GBIF, the median location uncertainty was about 100 m. For both platforms, half of the observations had a reported uncertainty between 10 m and 1,000 m (Supplementary Figure 5a). Flickr data showed a bias towards high precision; about 90% of its georeferenced images were recorded at city-level accuracy (level 11) or higher, and 30.1% were recorded at the highest precision of street-level (level 16) (Supplementary Figure 5b).

Georeferenced salmonid observations were densest in North America and Europe (Figure 4.3). iNaturalist had the widest global coverage, with dense clusters in North America and Europe, some observations in Russia and Japan, and a few records from South America, Australia, and New Zealand. GBIF data were densest in Europe, and Flickr data were concentrated in North America and Europe. We found few observations from the Southern Hemisphere on any platform.

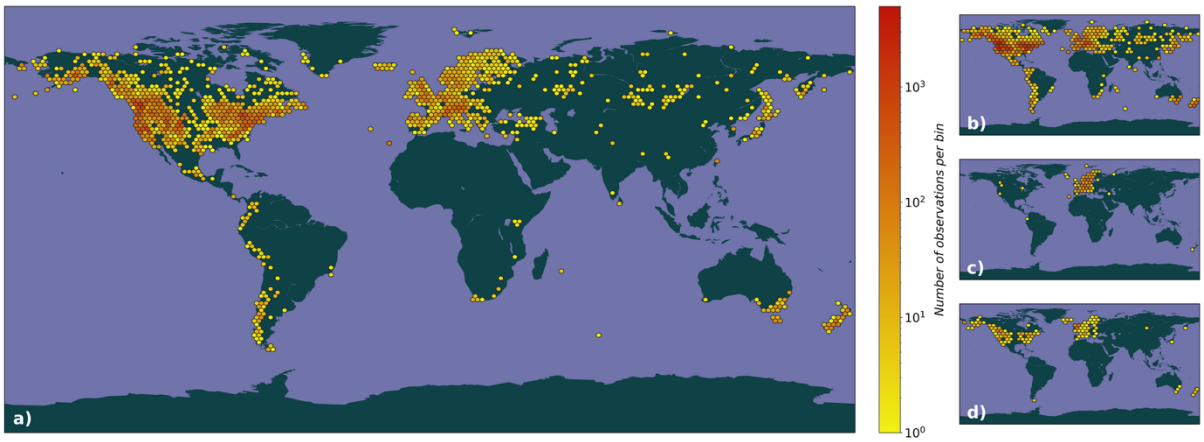


Figure 4.3. Global distribution of salmonid observations across image repositories. a) Combined observations from all sources, with hexagonal bins indicating observation density on a logarithmic scale from 1 (yellow) to >1000 (red) observations per bin. Individual distribution maps for b) iNaturalist, c) GBIF, and d) Flickr using the same binning and colour scheme.

The annual submission trends differed between platforms (Figure 4.4). GBIF had the earliest records, starting in the 1950s, and grew exponentially from around 2008. Flickr data showed a small peak around 1980, rose again in the mid-2000s, and then declined after 2010. iNaturalist showed rapid and sustained growth after its 2008 launch. Wikimedia Commons had fewer than ten submissions each year.

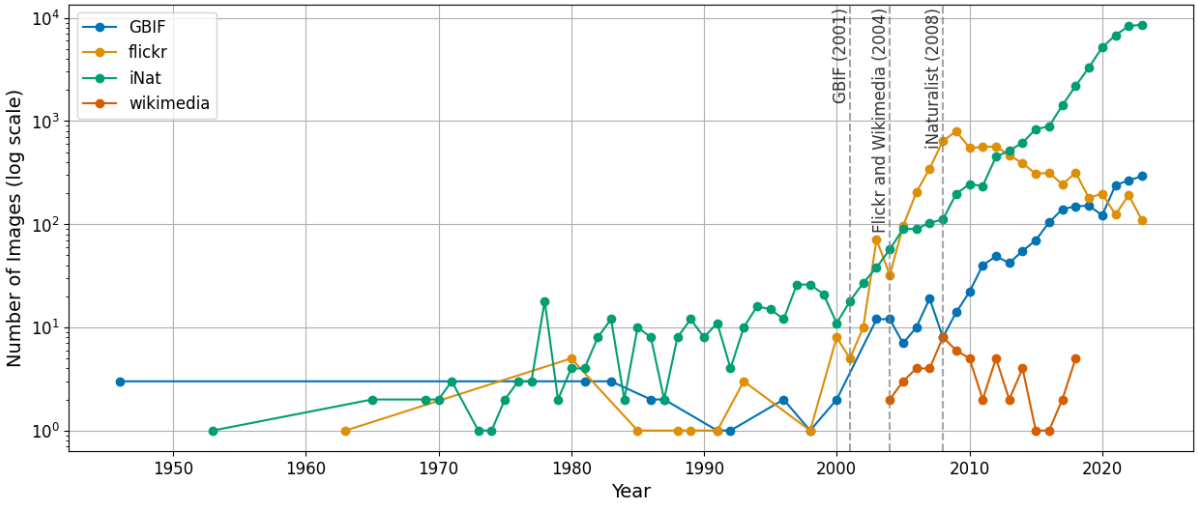


Figure 4.4. Temporal distribution of fish images across data sources from 1946 to 2023. The plot shows the number of images (log scale) contributed annually from each platform using their respective date fields: iNaturalist (observed_on), GBIF (eventDate), Flickr (datetaken), and Wikimedia (EXIF data). Vertical dashed lines indicate platform launch dates: GBIF (2001), Flickr and Wikimedia (2004), and iNaturalist (2008).

Seasonal patterns peaked in summer–autumn (June–October), which comprised 66.0% of observations, and the monthly distribution differed from uniform (=14373.43, df=11, p<0.001; Figure 4.5). Monthly patterns also differed slightly among platforms (=882.69, df=33, p<0.001; Cramér's V=0.077), with iNaturalist contributing the most observations in every month and Flickr showing a disproportionate January peak (standardised residual = 21.2).

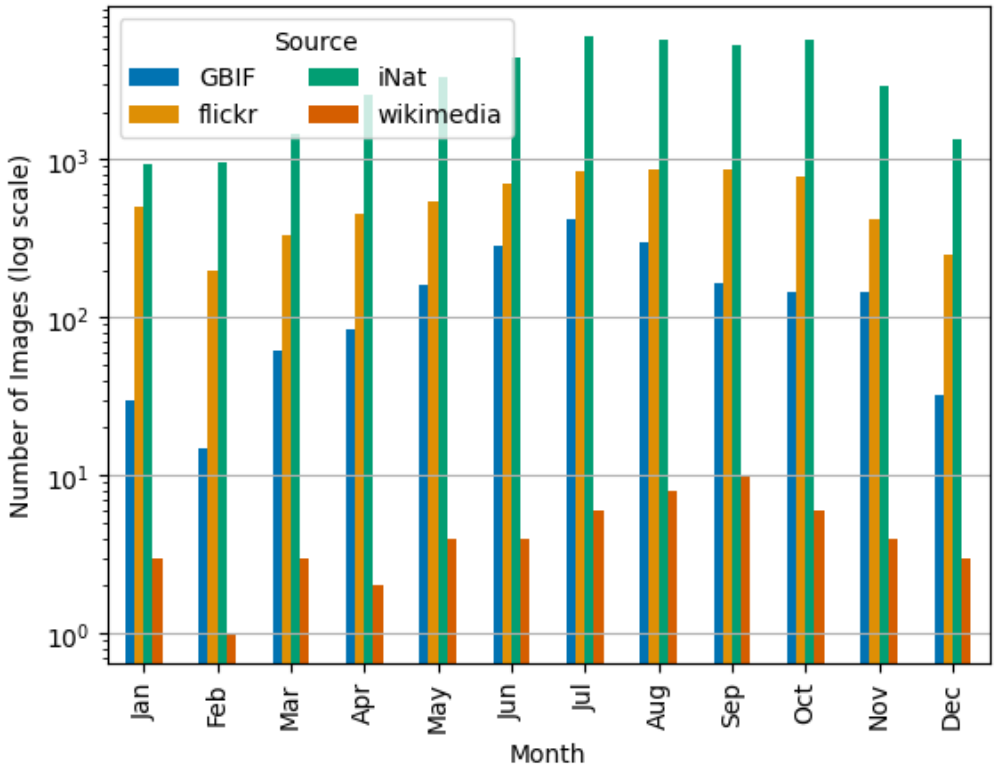


Figure 4.5. Monthly distribution of salmonid observations across image repositories (log scale).

4.3.3 Taxonomic Composition and Identification Patterns

The taxonomic resolution – the most specific level of identification for an observation – varied between the data sources. All salmonid observations on Wikimedia Commons were identified to species level. Species-level identification was also the most common resolution on iNaturalist (94%) and Flickr (91%). While most GBIF observations were also identified to species (72%), it had the widest range of other resolutions, with 21% of its records identified to subspecies and 4.4% to variety. A small number of records on Flickr could only be identified to a broader level, such as genus (6%) or subfamily (3.4%). Identifications that could not go beyond the family level were minimal on all platforms.

The hierarchical distribution of all salmonid observations shows that the subfamily Salmoninae is dominant (Figure 4.6). The most frequent species was *Oncorhynchus mykiss* (rainbow trout, 10,984 images), followed by *Salmo trutta* (brown trout, 9,573 images), and *Salvelinus fontinalis* (brook trout, 5,870 images). Other numerous

species included *Oncorhynchus tshawytscha* (Chinook salmon, 3,613 images) and *Oncorhynchus kisutch* (coho salmon, 2,724 images). The subfamilies Thymallinae and Coregoninae made up much smaller parts of the dataset. iNaturalist was the main source for most species.

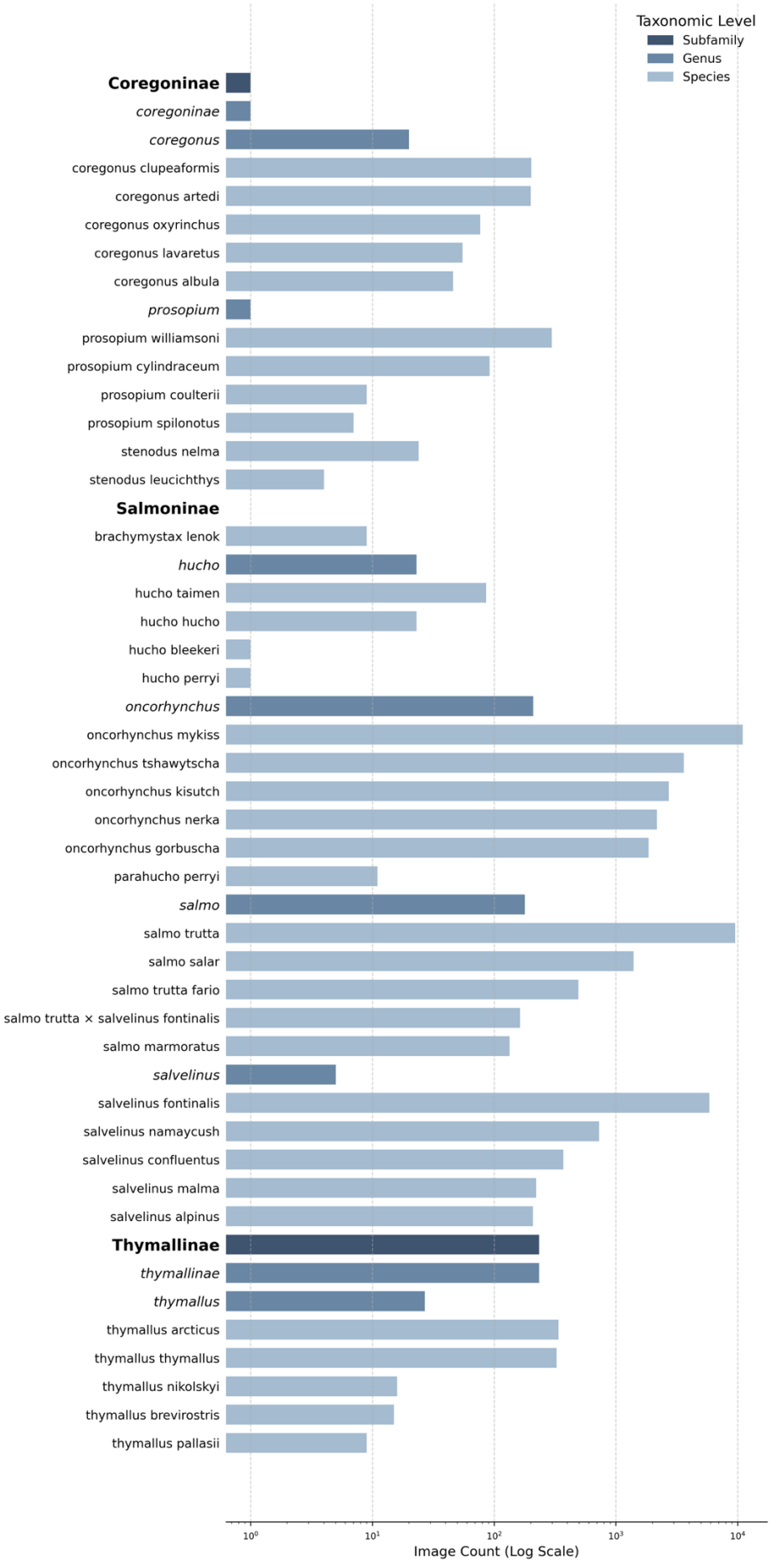

Taxonomic Level
Subfamily
Genus
Species
Coregoninae
coregoninae
coregonus
coregonus clupeaformis
coregonus artedi
coregonus oxyrinchus
coregonus lavaretus
coregonus albula
prosopium
prosopium williamsoni
prosopium cylindraceum
prosopium coulterii
prosopium spilonotus
stenodus nelma
stenodus leucichthys
Salmoninae
brachymystax lenok
hucho
hucho taimen
hucho hucho
hucho bleekeri
hucho perryi
oncorhynchus
oncorhynchus mykiss
oncorhynchus tshawytscha
oncorhynchus kisutch
oncorhynchus nerka
oncorhynchus gorbuscha
parahucho perryi
salmo
salmo trutta
salmo salar
salmo trutta fario
salmo trutta × salvelinus fontinalis
salmo marmoratus
salvelinus
salvelinus fontinalis
salvelinus namaycush
salvelinus confluentus
salvelinus malma
salvelinus alpinus
Thymallinae
thymallinae
thymallus
thymallus arcticus
thymallus thymallus
thymallus nikolskyi
thymallus brevirostris
thymallus pallasii
10^0
10^1
10^2
10^3
10^4
Image Count (Log Scale)

Figure 4.6. Hierarchical distribution of salmonid observations. The horizontal bar chart displays the taxonomic hierarchy, with genera indented under their respective subfamilies and species indented under their parent genus. The length of each bar represents the total number of images identified at that specific taxonomic level, plotted on a logarithmic scale to accommodate the wide range of counts. The colour of each bar corresponds to its taxonomic rank (Subfamily, Genus, or Species). For visual clarity, only the five most abundant species within each genus are displayed.

4.3.4 Image Content Analysis: Size and Health Indicators

We annotated a subset of 5,667 salmonid images. From this subset, we could not reliably assess 696 images because of poor quality or obstructions. Of the remaining 4,971 images, only 123 (2.5%) had standard size reference objects, which made systematic size estimation impractical.

For health assessment, 866 (17.4%) displayed visible signs of disease or physical damage (e.g., rashes, discolouration, parasites, growths, cuts, wounds, scarring), while 4,105 (82.6%) depicted apparently healthy salmonids. The prevalence of disease or damage indicators was largely consistent across platforms within the annotated subset: 18.0% for iNaturalist (755 of 4,203 images), 17.0% for GBIF (19 of 112 images), 16.7% for Wikimedia Commons (2 of 12 images), and 13.9% for Flickr (90 of 644 images).

4.4 Discussion

Our study used salmonids as a case study to assess how online image repositories can be used for ecological research. We showed that the choice of platform determines the quality and type of data available, so researchers must select platforms based on their specific needs. For studies needing rich metadata, iNaturalist and GBIF are the strongest sources. iNaturalist supplied the most verified salmonid images, and GBIF had complete temporal and spatial metadata for all its records. In contrast, Flickr gave many images initially but fewer verified salmonids and had limited location data. Wikimedia Commons made a very small contribution.

Our work supports earlier findings that combining data from multiple sources increases the number of available images (Heberling et al. 2021). However, we show that the usefulness of each platform depends on the research question and the required data quality. These repositories offer a large and growing visual dataset with wide spatial and temporal scope, showing they can supplement traditional monitoring. The value of these images for more advanced uses is limited by platform-specific problems. The high number of 'All Rights Reserved' licenses, mostly on Flickr, limits data reuse. The lack of standard size references on all platforms makes size estimation impractical. In contrast, our finding of a consistent 14-18% of images showing visible disease or damage suggests an opportunity for wildlife health surveillance using the 'extended image' idea.

4.4.1 Interpreting the Dataset: Availability, Quality, and Usability

The purpose of each platform and its user community likely explain the differences in the data we found. The success of iNaturalist shows the power of dedicated citizen science platforms for gathering verifiable observations. This contrasts with the challenges posed by general-purpose platforms like Flickr and Wikimedia Commons, whose broader aims likely explain their lower yield of relevant, verifiable data. Curated databases like GBIF are valuable for collecting research-grade data from many sources. Although these online repositories provide large datasets, other valuable collections may exist in less accessible places like local archives or private collections. Making these hidden datasets usable requires targeted outreach and adherence to data management principles that make them Findable, Accessible, Interoperable, and Reusable (FAIR) (Wilkinson et al. 2016; Reyserhove et al. 2020).

The image quality showed modest platform-specific tendencies. GBIF and Wikimedia generally scored slightly better on perceptual metrics (lower BRISQUE/NIQE), and GBIF tended to have sharper images on the Laplacian-variance metric (higher values), but platform effects were small (<0.01) with substantial overlap among sources. Consequently, image-level filtering is likely more useful than relying on platform-wide averages for downstream analyses.

Licensing restrictions are prohibitive to data reuse for the extended image concept. The high proportion of All Rights Reserved images (21.5%), mostly on Flickr, are barriers to curating

and sharing datasets or the products of computational analyses, particularly AI models. These activities involve copyright and the creation of derivative works (e.g. Vendrow et al. 2024), limiting potential for an image to be truly 'extended' through reprocessing. Researchers using online imagery should therefore evaluate licensing terms at project commencement, aligning with best research practices. This is particularly true for machine learning and data sharing, where attention to image provenance, legal frameworks, and community standards for data stewardship are essential (Wilkinson et al. 2016; Gebru et al. 2021; Hemphill et al. 2022; London School of Economics 2025).

4.4.2 Spatial and Temporal Dynamics of Online Salmonid Imagery: Opportunities and Biases

The opportunities and biases in the spatial data highlight the need for careful filtering and interpretation. The median location uncertainty for iNaturalist and GBIF records is about 100 m, which suggests the data can be used for fine-scale distribution modelling. The apparent high precision of Flickr data is promising, but its accuracy scale defaults to 'street-level' when users give no information. This default means many records may not reflect user-confirmed precision, misrepresenting the true accuracy. This issue, combined with the wide uncertainty range in iNaturalist and the lack of explicit distance values for Flickr's scale, means the data must be filtered and interpreted carefully for precise spatial work. The small number of georeferenced Flickr images further limits the use of this large image pool for spatial analysis.

The geographical concentration of images in North America and Europe matches the known native distribution of salmonids (Araneda et al. 2008). iNaturalist had the most extensive global coverage, but observations from the Southern Hemisphere were sparse across all platforms. This is consistent with salmonids being non-native to most of this region, with some introduced populations in New Zealand, Australia, and South America (Klemetsen et al. 2003). The dataset therefore captures native areas well but may underrepresent introduced ranges or show gaps in citizen science monitoring. This shows the importance of comparing image-derived distributions to established biogeographical knowledge (Klemetsen et al. 2003; Araneda et al. 2008) and accounting for sampling effort when inferring species distributions.

The annual increase in submissions, particularly from iNaturalist, shows a growing data stream that can be used for long-term studies, such as tracking changes in seasonal events. The seasonal peaks in summer and autumn align with expected increases in fishing activity and salmonid visibility (e.g., spawning runs) in the Northern Hemisphere. The unusual 1980 peak in Flickr data is also notable. Our inspection confirmed it was caused by incorrect metadata from two users who likely uploaded old photos with the wrong date, as Flickr launched in 2004. This shows a problem for temporal analysis, as the 'date taken' field may not always be the true observation date and needs to be checked.

4.4.3 Interpreting Taxonomic Patterns in the Salmonid Image Dataset

The high number of species-level identifications across platforms supports using online images for species occurrence data (Theobald et al. 2015; Heberling et al. 2021; Szentivanyi and Vincze 2022). The origin and reliability of these identifications vary, however. For Flickr and Wikimedia Commons, we inferred species from the original search terms (e.g., 'Rainbow Trout'). While we confirmed these images showed salmonids, the species label (e.g., *Oncorhynchus mykiss*) was not validated with the same rigour as a 'research grade' iNaturalist observation. This uncertainty, along with challenges like poor image quality and taxonomic ambiguity, can make species identification difficult even for experts (Marshall et al. 2020). These data therefore need careful interpretation based on their source, particularly for detailed ecological analysis.

The dominance of *Oncorhynchus mykiss* and other popular angling targets like Pacific and Atlantic salmon (Figure 4.6) likely reflects their widespread distribution, popularity with anglers, and how easy they are to identify from photos. This suggests human interests and likelihood of observation, not systematic surveying, shape the dataset. The lower representation of groups like Thymallinae and Coregoninae might also reflect these biases.

Despite these variations in taxonomic certainty, the dataset is valuable. Even images with less certain species labels can be used for distribution maps at the family level. They are also useful for training AI models when resources to re-check every image are limited (see Heberling et al. 2021). Such models can be designed to learn from datasets with 'noisy' labels, a term for labels that may be incorrect. The key is to use the strengths of each data source for the specific research question. This could mean using iNaturalist's community-

vetted taxonomy, GBIF's taxonomic depth, or the visual diversity from Flickr and Wikimedia for AI development.

4.4.4 Image Content Analysis: Size and Health Indicators

It was impractical to estimate fish size systematically because only 2.5% of images had standard size references. Assessing health indicators was a more practical approach. Our finding that 17.4% of suitable images showed visible signs of disease or damage confirms that these platforms can be used as an opportunistic data source for fish health surveillance. The prevalence of these signs was consistent across platforms, from 13.9% on Flickr to 18.0% on iNaturalist. This suggests the platforms capture these conditions uniformly, regardless of their main purpose or user base. The types of indicators we saw are also similar to those used in traditional fisheries health assessments, which implies these images could supplement existing monitoring programmes. Building on this finding, the data collected in this study on healthy salmonids and those with visible signs of disease formed the basis of a subsequent study that successfully trained a computer vision model for the automated classification of saprolegniasis-like infections (Olsen et al. submitted; Chapter 5).

However, using online images for scientific analysis faces challenges from image alterations, such as the common use of watermarks. An emerging problem is AI-driven image editing. Features like those in the Google Pixel 8 that automatically alter images (Berrada 2023), undermines scientific integrity if the alterations are not transparently documented. While metadata standards like those from the International Press Telecommunications Council (IPTC) can record alterations, the consistent documentation of AI-driven changes is not yet common (IPTC Photo Metadata Working Group 2024). This is a gap that initiatives like the Coalition for Content Provenance and Authenticity (C2PA) are designed to address (Coalition for Content Provenance and Authenticity 2025).

4.4.5 Limitations and Biases

While the 'extended image' concept offers opportunities, we must acknowledge several limitations. These reflect challenges found in the extended specimen and digital extended specimen literature, such as problems with data quality, standardisation, and integration (Lendemer et al. 2020; Hardisty et al. 2022). Common challenges for online images include geographical bias (e.g. Durso et al. 2021), limited taxonomic coverage, variable image

quality, and incomplete metadata. Platform-specific biases from search algorithms or user behaviour can also affect how representative the data is (Jarić et al. 2020; Vardi et al. 2024). The Digital Extended Specimen framework (Hardisty et al. 2022) highlights further needs for persistent identifiers, machine-actionable data, and clear provenance. These limitations, while significant, can be addressed through methodological rigour, careful consideration of sampling biases, and the adoption of best practices for data collection and analysis (Callaghan et al. 2019; Di Cecco et al. 2021).

4.4.6 Future Research and Recommendations

Future work should use all available image data. Images with incomplete metadata, like those from Flickr or Wikimedia Commons that lack precise locations, or those with unverified taxonomic labels, should not be dismissed. These images are useful for training robust AI and machine learning models. For example, such models can be built for general salmonid detection using foundational models (Bommasani et al. 2021) or for species identification systems that tolerate noisy labels (Sun et al. 2021). They could also be used for visual assessment of health indicators across broader taxonomic groups (e.g. Siachos et al. 2024).

Several strategies could also improve online repositories for science. Standardising metadata and image quality metrics would help when combining data from different platforms. Automated verification tools could help process large numbers of images, though they need carefully curated training data that accounts for platform differences. Adopting open licensing policies across all platforms would improve research access and reproducibility. Finally, engaging with contributor communities, such as anglers and photographers, to teach best practices for taking scientifically useful images could improve the quality of future data.

CHAPTER 5 :

COMPUTER VISION FOR INFECTIOUS DISEASE SURVEILLANCE; *SAPROLEGNIA* SPP. IN SALMONIDS

Agnethe S. Olsen[a], Paul L. Rosin[b], Christopher B. Jones[b], Jo Cable[a], Sarah E. Perkins[a]

[a]School of Biosciences and Water Research Institute, Cardiff University, Cardiff, CF10 3AX, United Kingdom

[b]School of Computer Science and Informatics, Cardiff University, Cardiff, CF24 4AG, United Kingdom

Abstract

Effective disease surveillance in wild fish populations is essential for food security and biodiversity conservation, but data acquisition can be limited by *ad hoc* reporting and resource-intensive laboratory diagnostics, limiting the spatio-temporal scope. We developed and evaluated a computer vision pipeline to detect saprolegniasis-like infections, a devastating disease affecting wild and farmed salmonids that manifests as visible signs on the fish.

Compiling a dataset of 4,526 images (494 infected, 4,032 healthy) from citizen science platforms and stakeholders, we used data augmentation to address the significant class imbalance. We then fine-tuned and compared four pre-trained convolutional neural network architectures (EfficientNetV2S, EfficientNetV2B0, ResNet50, and MobileNetV3S), chosen to represent a range of standard and efficient models, to classify healthy versus infected fish across datasets of varying host taxonomic specificity.

The EfficientNetV2S model achieved the highest performance on a *Salmo* genus-specific evaluation dataset, with a mean recall (proportion of infected fish images correctly identified) of 0.918 (± 0.038) and precision (proportion of correctly identified infected fish among all fish identified as infected) of 0.862 (± 0.056). Performance differed between *Salmo* and *Oncorhynchus* evaluation sets and depended on the training data strategy, with genus-specific training performing best on *Salmo* images, while broader training improved recall on *Oncorhynchus* images at the expense of precision. Despite challenges including variable image quality, water surface reflections, and inherent class imbalance, these results show computer vision can support large-scale disease surveillance in wild fish populations.

Computer vision-based surveillance could enable earlier outbreak detection and targeted diagnostics, improving freshwater ecosystem health management. While successful implementation hinges on acquiring sufficient high-quality imagery, this study highlights the potential of applying tailored Artificial Intelligence tools for monitoring visually detectable diseases across diverse wildlife species.

5.1 Introduction

The emergent field of iEcology (internet ecology) frequently uses image repositories to generate insights into species distribution and occurrence, biogeographical patterns, behaviour, species interactions, habitat use, and the impact of human activities on wildlife (Weinstein 2018; Jarić et al. 2020; Tuia et al. 2022). Its application to wildlife disease surveillance, however, remains sparse, despite the abundance of available data (Edwards et al. 2021). While examples are emerging, such as automated classification of Devil Facial Tumour Disease (DFTD) in Tasmanian devils (*Sarcophilus harrisii*) (Nurçin et al. 2024) and lesions in bottlenose dolphins (*Tursiops erebennus*) (Murphy et al. 2025), the field is still under-explored. Computer vision offers a promising tool for automating disease detection from images where diseases have visible signs. Freshwater wild fish are an excellent test bed for disease detection because they are a significant source of diseases (Shinn et al. 2014), including zoonoses (Gauthier 2015), they represent some of the most threatened vertebrates on the planet (Collen et al. 2014; Shinn et al. 2014; Dias et al. 2017), and disease control costs billions (Shinn et al. 2014). Many fish diseases present visible signs, making them well-suited for image-based detection. Furthermore, fishers, especially recreational anglers, have a culture of taking and sharing images online of their catch, therefore in theory, both visible signs and the images from which to observe them exist. The advantages of image-based disease detection over traditional surveillance include the ability to rapidly screen thousands of images in a non-invasive manner.

Computer vision applications have demonstrated promising results for disease detection across domesticated and livestock species. Deep learning models have shown success in diagnosing ocular surface diseases in domestic dogs and cats (Nam and Dong 2023) and detecting skin conditions like pododermatitis and neoplasia in dogs (Smith et al. 2024). Computer vision has been used to detect early signs of respiratory diseases in pigs via

changes in temperature using thermal imagery (Jorquera-Chavez et al. 2020). In cattle, computer vision has been employed for tick detection and identification (Barbedo et al. 2017; Luo et al. 2022) and for real-time detection and scoring of digital dermatitis (Aravamuthan et al. 2024).

While these advancements showcase the potential of computer vision in veterinary medicine, research on wild species is limited. Previous studies in fish have primarily focused on aquaculture settings using small datasets, often with limited information on the species involved (Malik et al. 2017; Hasan et al. 2022; Mia et al. 2022; Yasruddin et al. 2022; Rachman et al. 2023; Vijayalakshmi et al. 2023; Biswas et al. 2024; Kumaar et al. 2024; Maruf et al. 2024). Ahmed et al. (2022), for example, classified 'salmon disease' in 266 images of salmon (83 healthy, 183 infected) with an accuracy of 91.4% using traditional computer vision methods, but the disease and salmon species were not specified. Gupta et al. (2022) achieved an accuracy of 96.7% using convolutional neural networks to classify 3,289 salmon images (augmented based on an initial dataset of 68 healthy, 71 wounded, 70 with fish-lice). Importantly, inclusion of augmented images during model validation and testing may influence evaluation metrics (Huang and Khabusi 2023; Rachman et al. 2023; Biswas et al. 2024; Maruf et al. 2024).

There is a paucity of surveillance of fish diseases, such as saprolegniasis, caused by the oomycete *Saprolegnia parasitica*, which kills 1 in 10 farmed salmon (Dias et al. 2017). This disease, characterised by fungal-like white growths on the fish's body, head, and fins, has no effective treatment. It causes significant morbidity and mortality in wild fish populations (van West and Beakes 2014; Derevnina et al. 2016; Matthews 2019; Matthews et al. 2021) and can infect other aquatic species (Costa and Lopes 2022). With widescale fish mortality and inter-specific transmission, early detection of this disease is critical to assess risk to aquatic species, and to help identify drivers of outbreaks (MacAulay et al. 2022). While regular disease surveillance occurs in commercial fisheries, outbreaks in wild fish are often detected by *ad hoc* reporting of diseased/dead fish to the relevant fisheries authorities (e.g. Fish Health Inspectorate in the UK). Once an outbreak is recognised as such, identification of *Saprolegnia* spp. may follow with direct sampling of animals (Tandel et al. 2021) or water (Pavić et al. 2022) using molecular methods or culturing. While these methods are highly sensitive, they are time-consuming and damage to fish stocks has usually occurred by the

time the pathogen is identified. Because *Saprolegnia* spp. cause visible signs of infection there is potential for use of image-based disease surveillance. While visible signs alone cannot confirm disease, these data could support large-scale surveillance and identify areas for targeted investigation.

Here, we establish a computer vision pipeline to detect common infectious diseases (*Saprolegnia* spp.) in wild salmonids. We use standard methods (Jarić et al. 2020; Edwards et al. 2021) to collate a large database of images of wild fish, and assess the accuracy with which computer vision classifiers detect infectious disease. Our ultimate aim is to assess if we can move current practices from intermittent, episodic reports of disease towards detailed real-time monitoring of wild freshwater fish. Doing so would provide a step towards collating spatiotemporal information on disease and provide a framework from which we can expand to other host-parasite systems. Our proposed pipeline offers a step-change in monitoring infectious diseases, providing a technology-led framework for understanding disease dynamics in wild fish and other species.

5.2 Methods

5.2.1 Data

Images of salmonids were acquired from photo-sharing websites, Flickr (www.flickr.com), iNaturalist (www.iNaturalist.org), GBIF (www.GBIF.org) and Wikimedia Commons, between December 2023 and February 2024. These data constitute 'passive citizen science'; images submitted without scientific intent[1] that nonetheless could contain important ecological data (Edwards et al. 2021). These sites were accessed using Application Programming Interfaces (APIs) that allow for keyword or taxonomic-level searches to download images and associated metadata.

Image Acquisition

We focused our work on salmonids as the taxa most prone to saprolegniasis (Vieira da Silva do Nascimento et al. 2020). GBIF and iNaturalist can be searched taxonomically for 'Salmonidae', returning observations linked to any subfamily, genus or species within this

[1] Upon manual review of the images, titles and descriptions, it is clear that most images were taken by anglers.

family. Flickr and Wikimedia Commons allow for keyword searches in text fields such as titles and descriptions. To compile keywords, a comprehensive list of scientific and common salmonid names was created by integrating subfamily, genus, and species data from FishBase (https://www.fishbase.se/search.php), FishTreeOfLife (https://fishtreeoflife.org/api/taxonomy/family/Salmonidae.json), and NCBI (https://www.ncbi.nlm.nih.gov/Taxonomy/Browser/wwwtax.cgi). This list was enriched with additional species and English common names from FishBase and iNaturalist (see Supplementary Table 6 for a full list of taxa). This resulted in a final list of three subfamilies, 11 genera, 387 species and two hybrids.

We searched GBIF occurrence data using the following filters: "BasisOfRecord is Human Observation", "MediaType is Image", "OccurenceStatus is Present", "TaxonKey is Salmonidae". Of the returned occurrences, 94.5% were duplicated from iNaturalist and these duplicates were excluded. We downloaded the Darwin Core Archive resulting from our search and used the gbif-dl (v0.1.1) package in Python 3.9.18 to extract the corresponding URLs and download the images.

For iNaturalist images we first downloaded the metadata files (observations, observers, photos and taxa) from the 'iNaturalist Licensed Observation Images' open dataset (https://registry.opendata.aws/inaturalist-open-data) using the AWS Command Line Interface (CLI). We then used Python 3.11.7 and the pandas (v1.5.3) package to link taxa information to observation, observer and photo metadata, allowing us to filter to images identified as the Family 'Salmonidae' (taxon_id=47520) or lower associated taxonomic levels. We used the multiprocess (v0.70.15) package to download the images. To ensure comprehensive coverage, we incorporated iNaturalist observations regardless of their quality grade.

We queried the Flickr API for images with tags, descriptions or titles containing the terms in Supplementary Table 6 using the flickrapi (v2.4.0) package in python 3.11.7. We tested the search terms in the Flickr user interface first and excluded terms that returned a very high proportion of irrelevant images, such as the common name 'salmon' which returns over 400,000 images, mainly of salmon prepared for food. SHA-256 (Secure Hash Algorithm) from

hashlib in the Python Standard Library of python 3.11.7 was used to create unique ids for each image to locate and remove duplicate downloads.

To acquire images from Wikimedia Commons we followed the process implemented by Marshall et al. (2020), amending their R script (SuppCode2_Wikimedia_query.R from zenodo.org/records/4010155) to query for the terms listed in Supplementary Table 6. Following removal of duplicate image URLs, the images were downloaded using urllib3 (v1.26.16) in Python 3.11.7.

In total, before 'ground-truthing' to check for salmonids, 69,158 photographs were collated across online sources based on keyword or taxonomic-level searches: 49,057 from iNaturalist, 1,843 from GBIF, 19,610 from Flickr and 183 from Wikimedia Commons.

Ground Truth - Images of Healthy and Diseased Salmonids

The Wikimedia Commons and Flickr image datasets contained a high proportion of irrelevant images, as they were downloaded based on keywords (Supplementary Table 6) that often had multiple meanings. For example, a Grayling is also a moth species. To address this issue two labellers manually screened the dataset for the presence of salmonids (Figure 5.1), on Labelbox (Labelbox. 2025), leaving 91 relevant images from Wikimedia Commons and 6,869 from Flickr which were added to the existing collection of 49,057 images from iNaturalist and 1,843 images from GBIF.

Due to the time-consuming nature of expert annotation and the requirement of specific expertise to identify *Saprolegnia* spp., two of the authors examined a subset of approximately 10% (5,667) of the salmonid images, stratified by data source and taxonomic classification, for visible signs of disease, outlined in Figure 5.1. This subset provided a manageable starting point for the intensive labelling process, while still yielding a dataset large enough to develop and validate our proof-of-concept pipeline. Due to poor image quality or obstructions making it difficult to assess fish health, 696 images were excluded. From this initial evaluation, 4,105 images were identified as containing healthy salmonids, making up the first class in our binary classification problem. The remaining 866 images were reviewed for visible signs of *Saprolegnia* spp. Where more than one salmonid was visible in an image, the image would be labelled as part of the '*Saprolegnia* spp.' class if at

least one of them displayed visible signs of the disease. This process identified 217 potential infections in salmonids, making up the second class in the classification problem.

The '*Saprolegnia* spp.' class was supplemented following the creation of ground truth labels by searching the iNaturalist and Flickr APIs using the keyword '*Saprolegnia*' ('Additional *Saprolegnia* spp. images' in Figure 5.1). The Flickr search returned 33 images with nine new images showing salmonids with visible signs of *Saprolegnia* spp. infection while iNaturalist returned 42 images of which 41 showed signs of saprolegniasis. Additionally, we incorporated 198 images provided by the Environment Agency's National Fisheries Laboratory (EA), 120 images uploaded to the Fisheries Management Scotland (FMS) app for fish disease (https://fms.scot/fish-health-and-disease), and 55 provided by the Fish Pathobiology and Immunology Laboratory at Michigan State University. These photographs were assumed, not verified, to show fish with *Saprolegnia* spp. based on visible signs. The final count of images with *Saprolegnia* spp. infection was 630.

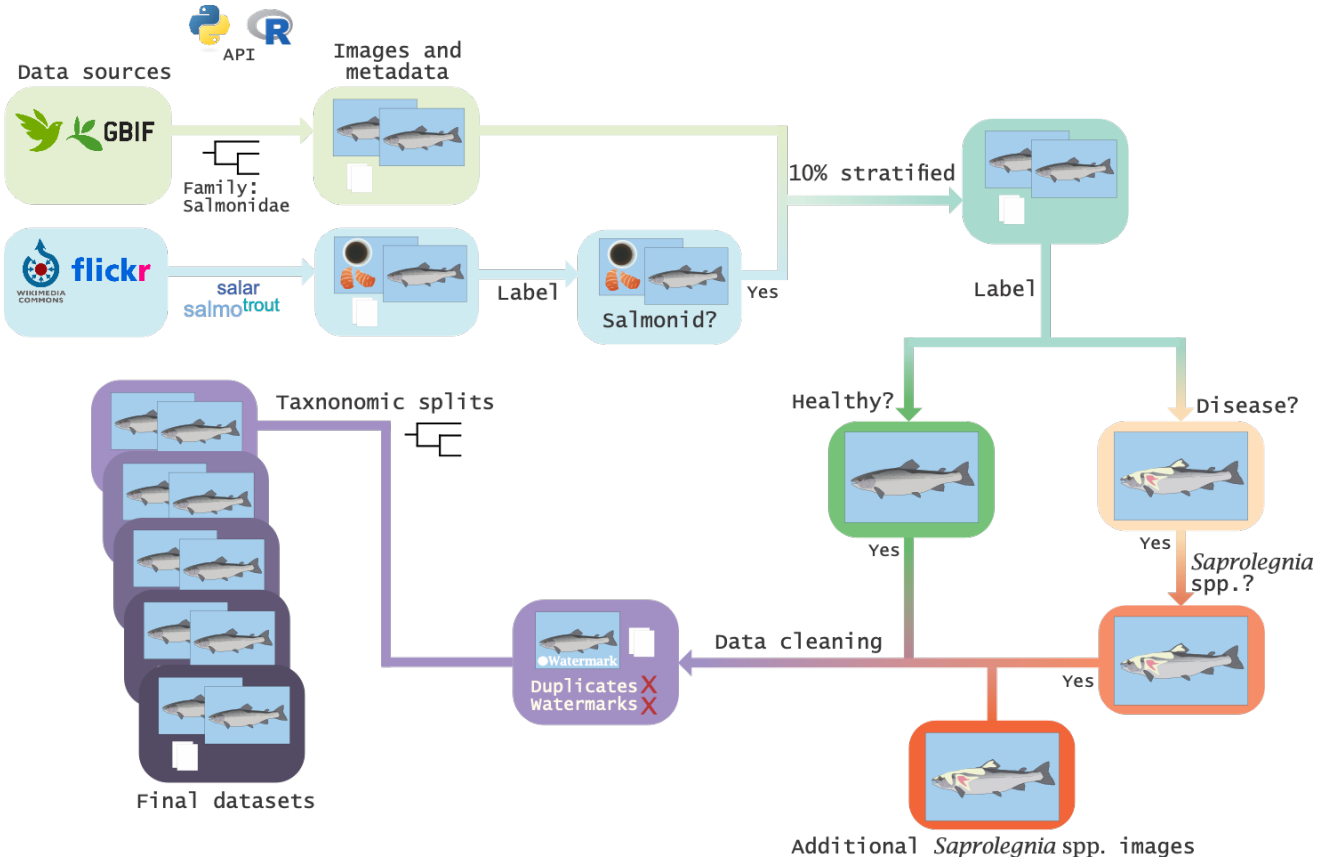


Figure 5.1. Data pipeline for collating images and identifying disease (*Saprolegnia* spp.) in salmonids. Images and metadata were downloaded from online sources using API searches based on taxonomy or search terms and false positives removed. A 10% subset, stratified by data source and taxonomic classification, of the data was labelled as 'healthy' or 'disease' and the 'disease' class was subsequently screened for visible signs of *Saprolegnia* spp. Images with visible signs of *Saprolegnia* spp. infections were added from additional sources (e.g. Environment Agency). Following cleaning, the data were split into tiered datasets based on taxonomic classification and the number of available images in the 'healthy' and '*Saprolegnia* spp.' classes per taxon.

Image Cleaning

All images were manually cropped to remove obvious watermarks, time stamps, duplicates or borders. If this was not possible the image was excluded. The final dataset consisted of 4,032 images showing healthy salmonids, hereafter referred to as the 'healthy' class, and 494 images showing salmonids with visible signs of *Saprolegnia* spp. infection, hereafter referred to as the '*Saprolegnia* spp.' class (Table 5.1).

Table 5.1. The number of salmonid images classed as 'healthy' and '*Saprolegnia*' spp. from each image source.

Source	**'Healthy' count**	***'Saprolegnia* spp.' count**
iNaturalist	3,374	211
Flickr	556	22
GBIF	92	2
Wikimedia Commons	10	0
Environment Agency	0	130
Fisheries Management Scotland	0	98
Michigan State University	0	31

Total	4,032	494

Metadata

Metadata associated with images was used to split the data into training and validation partitions. iNaturalist, Flickr, FMS and Immunology Laboratory at Michigan State University images were accompanied by detailed metadata, including user information, date, and location. While Wikimedia Commons images were downloaded with user and date information, location data was unavailable. EA images sourced from National Fisheries Laboratory (NFL) image archives, encompassing those taken by EA field officers, NFL employees, and anglers, included date and camera specifications as a proxy for user information in the EXIF data. Although location information was missing for the EA images, we knew that all were captured within England and Wales.

Dataset strategies

The 'healthy' and '*Saprolegnia* spp.' dataset comprised images of salmonids classified to different taxonomic levels, ranging from family to species. A total of 13 taxa had images of both healthy and infected classes (Table 5.2) with the remaining taxa having images only in one of the classes. To address the imbalance in *Saprolegnia* spp. infected and healthy images across different taxonomic groups, we implemented a multi-tiered data preparation strategy, based on taxonomic specificity and the number of available images for each taxonomic classification.

Table 5.2. Count and ratio of images across host taxa for Healthy and *Saprolegnia* spp. classes. Ratio represents the number of *Saprolegnia* spp. images divided by the number of Healthy images. Only taxa with images in both classes are included. See Supplementary Table 6 for a full list of taxa.

Host taxa	**Healthy**	***Saprolegnia* spp.**	**Ratio**
Salmo spp.	43	135	3.140
Salmo salar	91	110	1.209

Host taxa	Healthy	*Saprolegnia* spp.	Ratio
Oncorhynchus tshawytscha	180	54	0.300
Oncorhynchus spp.	181	45	0.249
Oncorhynchus keta	50	45	0.900
Oncorhynchus gorbuscha	85	27	0.318
Oncorhynchus nerka	132	25	0.189
Oncorhynchus mykiss	936	24	0.026
Oncorhynchus kisutch	214	11	0.051
Salmo trutta	836	10	0.012
Salvelinus fontinalis	529	2	0.004
Thymallus thymallus	44	2	0.045
Prosopium williamsoni	24	1	0.042

To address the imbalance in *Saprolegnia* spp. infected and healthy images across different taxonomic groups, we implemented a multi-tiered data preparation strategy. The full dataset, referred to as 'All photographs', included all 4,032 'healthy' and 494 '*Saprolegnia* spp.' images from 68 different taxa. We then created several tiered datasets (Table 5.3):

- 'Taxa present in both classes': This dataset included all images from taxa with images present in both the 'healthy' and '*Saprolegnia* spp.' classes (Supplementary Table 6).
- 'Taxa with ≥ 10 photographs in both classes': This dataset restricted the inclusion criteria to taxa with a minimum of 10 images in both classes. The 10-image threshold was chosen to ensure a reasonable number of images for training, validation, and testing, while also maintaining a balance between data quantity and quality.
- Genus specific datasets: To account for potential variations within specific genera, we created separate datasets for species within *Oncorhynchus* and *Salmo*, each with a minimum of 10 images per class, '*Oncorhynchus*' with ≥ 10 photographs in both classes' and '*Salmo* with ≥ 10 photographs in both classes'.

Both data quality and data quantity are important factors for model performance. While a larger, more diverse dataset can improve model generalisation, addressing potential class imbalance and intra-class variability is crucial. In our case, a larger dataset might introduce samples that, while belonging to the same class, exhibit significant visual differences. Conversely, a smaller, more focused dataset may lead to overfitting where the model starts to 'memorise' the training data rather than learning real patterns. Our strategy aimed to balance these trade-offs by focusing on taxa with adequate representation in both classes and by creating genus-specific datasets to capture intra-genus variations, and therefore our tiered datasets included a mixture of all photographs and taxa within and across broad taxa (Table 5.3). It was not practical to look at one species in isolation in this study, as the datasets would have been too small to be practically split into adequate training and validation sets.

Table 5.3. Count and ratio of images for the '*Saprolegnia* spp.' and 'Healthy' classes in the tiered datasets. The class ratio is calculated as the number of '*Saprolegnia* spp.' images divided by the number of 'Healthy' images.

Dataset	***Saprolegnia* spp.**	**Healthy**	**Class ratio**
All photographs	494	4,032	0.123
Taxa present in both classes	491	3,345	0.147
Taxa with 10 photographs in both classes	486	2,748	0.177
Oncorhynchus, 10 photographs in both classes	231	1,778	0.130
Salmo, 10 photographs in both classes	255	970	0.263

Dataset Splits

The limited sample size of small validation sets can introduce high variance in performance metrics, as the choice of validation samples can substantially impact the evaluation results (Chollet 2021). The best practice with small datasets is to use k-fold cross-validation (Chollet

2021). This technique divides the data into k 'folds' (subsets), using each fold once as a validation set while training on the remaining folds. We implemented this using the scikit-learn package StratifiedGroupKFold function, with k=5 folds. To ensure robust evaluation, we leveraged the metadata associated with each photograph to stratify by location and taxonomic classification, and group by user information (for example, username for iNaturalist submitted images) to prevent data leakage from user-specific patterns. These patterns, which persist even after removing obvious duplicates, can include consistent camera artifacts (such as sensor noise or colour profiles) or a characteristic photographic style. If these user-specific signatures were present in both the training and validation sets, the model's performance could be artificially inflated by learning to identify the photographer rather than the disease. This grouping ensured that photographs from the same user were kept within the same fold, to avoid data leakage from the training to validation sets. This approach was crucial, as the smallest tiered dataset contained only 255 images in the '*Saprolegnia* spp.' class, with some users contributing as many as 31 of these images. By using stratified group k-fold cross validation, we aimed to assess the stability of each model across different data splits. The reported metrics for each model represent the average performance across the 5 folds.

5.2.2 Image Classification

We selected four neural network architectures to compare a range of common and state-of-the-art approaches. We chose ResNet50 (He et al. 2016) because it is a widely recognised architecture used as a standard baseline for image classification, providing a robust point of comparison (Alom et al. 2019). ResNet-50 is a 50-layer residual network that uses skip connections between convolutional blocks to ease optimisation of deep architectures (He et al. 2016). We also included MobileNetV3S (Howard et al. 2019), a lightweight network based on depth wise separable convolutions and squeeze-and-excitation blocks, which is designed for high efficiency on less powerful devices. This is an important consideration for future work, where the model could be integrated into a mobile application for in-field analysis. Finally, we chose two models from the more modern EfficientNetV2 family (Tan and Le 2021), which are known for their high accuracy and computational efficiency compared to older models like ResNet50. This family of models uses compound scaling and Fused-MBConv blocks to achieve high accuracy with relatively few parameters (Tan and Le

2021). We specifically used two variants, EfficientNetV2B0 and EfficientNetV2S, to explore the trade-off between model size and performance on our dataset. EfficientNetV2B0 is the smaller, more efficient model, while the slightly larger EfficientNetV2S offers potentially higher accuracy at a greater computational cost. This selection allowed us to assess performance across different model backbones, balancing a classic baseline with modern, efficient alternatives.

Addressing Class Imbalance

Class imbalance is a common issue in machine learning, especially in image classification and wildlife classification tasks, such as species classification, where distributions are often long-tailed (Cunha et al. 2023). In this study, the datasets exhibit a clear imbalance, with a large majority of 'healthy' samples and a minority of '*Saprolegnia* spp.' samples; this is an expected feature of disease, reflecting the reality that for many diseases the number of healthy individuals vastly outnumber those with disease. This imbalance can lead to model bias towards the majority class. Because our goal is surveillance (minimising missed infections), we explicitly applied class-imbalance mitigation during training (oversampling and augmentation; see below) and report metrics that remain informative under imbalance.

Increasing the Sample Size

While oversampling (duplicating existing training samples) can potentially lead to overfitting (Alkhawaldeh et al. 2023), we mitigated this risk by subsequently applying data augmentation techniques to the expanded training dataset. Importantly, oversampling was applied only to the training partition within each cross-validation fold (not to validation data): after splitting images into a fold-specific training set and validation set, the minority class ('*Saprolegnia* spp.') in the training set was randomly sampled with replacement to increase its frequency. Data augmentation, a set of popular techniques to increase training data size, especially when samples are limited (Shorten and Khoshgoftaar 2019; Mumuni and Mumuni 2022), can create a more robust and varied dataset and enhance model generalization capabilities. RandAugment, a data augmentation technique that applies a combination of image transformations (Cubuk et al. 2020), was applied during training. RandAugment was implemented in Keras-CV with the number of augmentations per image, N = 3, and magnitude M = 0.5. Here, N controls how many random transformations are

applied to each training image, while M controls the overall strength of those transformations. Increasing N or M generally increases regularisation and robustness to nuisance variation, but if set too high can introduce unrealistic artefacts that may reduce performance; conversely, if set too low the regularisation effect may be insufficient. We used these standard settings as a practical balance and did not perform an exhaustive sensitivity analysis of N and M due to computational cost.

Transfer Learning

To leverage transfer learning, all models were instantiated with pre-trained weights from training on ImageNet (Krizhevsky et al. 2012). In practical terms, the weights are the learned numerical parameters of the network (convolutional filters and layer parameters) obtained by optimising model predictions to minimise a loss function on the ImageNet classification task. Initialising with these pre-trained parameters provides a strong starting point because early and mid-level layers typically capture generic visual features (e.g., edges, textures and shapes) that transfer well to many image tasks. The model can then be fine-tuned on this new, often smaller dataset, adapting these generic features to the specific visual patterns associated with, in our case, saprolegniasis-like infection in salmonid photographs.

Loss Function

The choice of loss function (a mathematical function that quantifies the difference between model predictions and actual observations) is crucial for addressing class imbalance. Focal Loss is a modified cross-entropy loss which down-weights the loss contributions of well-classified examples, allowing the model to focus on the more challenging minority class (Lin et al. 2020). We employed the Keras-CV implementation of Focal Loss with default parameters, which are effective in various computer vision classification tasks (Nemoto et al. 2018; Petmezas et al. 2022; Nie et al. 2023). We set the bias initialization of the final classification layer to $b=\log((1-\pi)/\pi)$, with $\pi=0.01$, as suggested by Lin et al. (2020), who show that this prevents large destabilizing loss values at the start of the training process.

Training Pipeline

The training pipeline was implemented in Python 3.10.13 using Keras 3 with JAX as the back end, and all models were trained on two Nvidia P100 GPUs. All four architectures shared the

same transfer-learning pipeline (Figure 5.2). Each cropped salmonid image was resized to the required input resolution (224x224 pixels for ResNet-50, MobileNetV3S and EfficientNetV2-B0, 300x300 pixels for EfficientNetV2S) and, during training, RandAugment (N = 3, M = 0.5) was applied to the training images. The augmented image was then passed through the chosen convolutional backbone initialised with ImageNet weights, followed by a custom classification head consisting of global average pooling, a fully connected layer (128 units for ResNet50 and MobileNetV3S, 256 for EfficientNetV2B0 and 512 for EfficientNetV2S), dropout with rate 0.2, and a final dense layer with two sigmoid-activated outputs ('healthy' and '*Saprolegnia* spp.'). The bias of the output layer was initialised to reflect a low prior probability of disease ($\pi = 0.01$, $\beta \approx -4.6$) for the focal loss, and the Adam optimiser was used to update model weights.

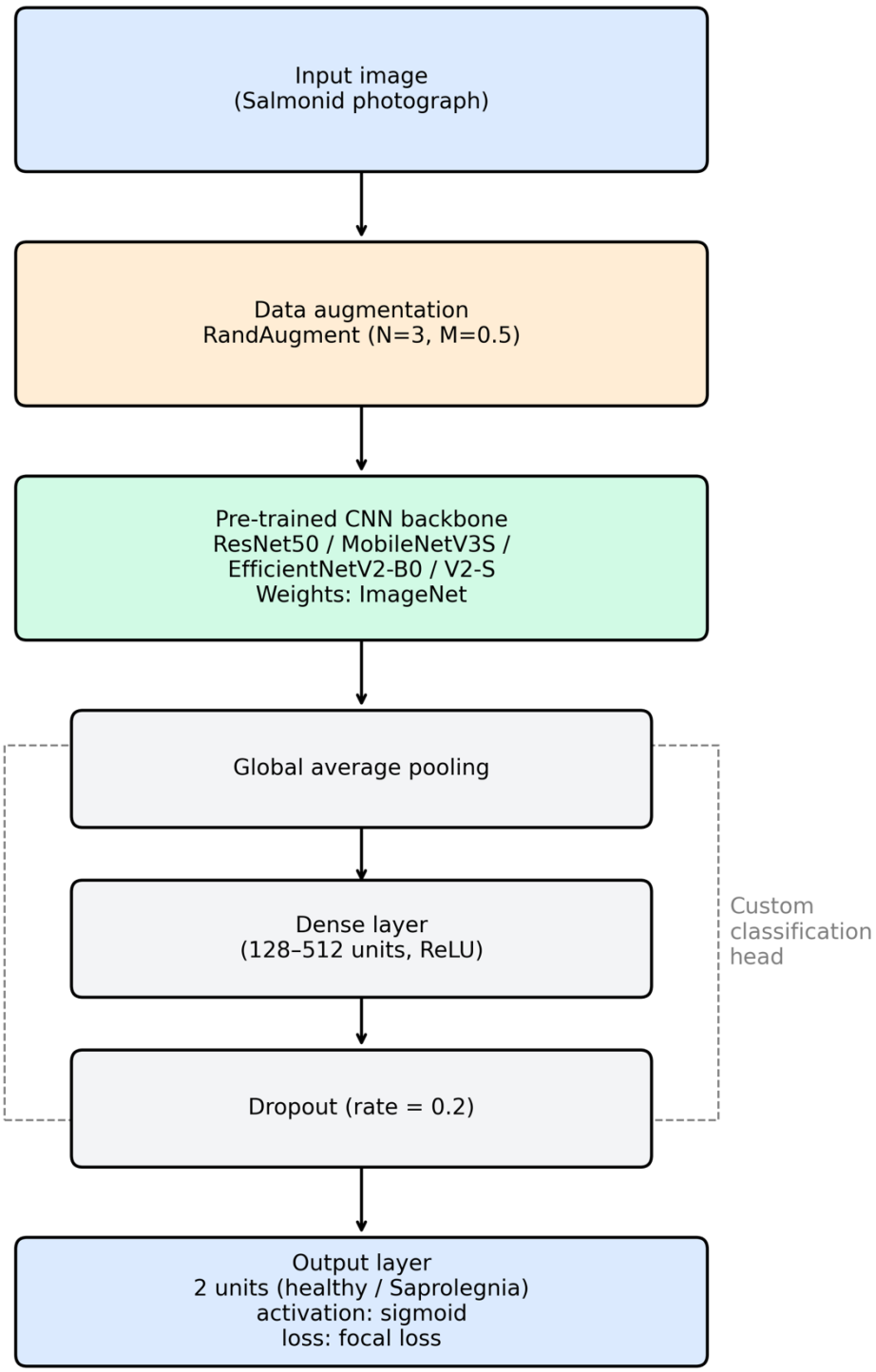


Figure 5.2. Transfer-learning pipeline used for training models on binary classification task. Each cropped salmonid photograph was resized and augmented using RandAugment (N = 3, M = 0.5) before being passed through a pre-trained convolutional backbone (ResNet50, MobileNetV3S, EfficientNetV2B0 or EfficientNetV2S, initialised with ImageNet weights). The backbone output was fed to a custom classification head consisting of global average

pooling, a dense layer (128–512 units, ReLU) and a dropout layer (rate 0.2), followed by a two-unit sigmoid output for the 'healthy' and '*Saprolegnia* spp.' classes trained with focal loss.

We used a typical transfer learning and fine-tuning approach (Chollet 2021). We instantiated the base model backbone with pre-trained, ImageNet (Krizhevsky et al. 2012), weights. First, the backbone layers were frozen and only the classification head was trained, using a learning rate of 0.001, a common and effective starting point for this training phase (Chollet 2021). If the validation loss did not improve for five epochs, the learning rate was reduced by a factor of 0.1. If the validation loss did not improve for 20 epochs, training in this frozen phase was stopped and the model weights were restored to those from the epoch with the lowest validation loss. In the second phase the last blocks of the backbone were unfrozen (as specified for each model in Table 5.4) and the whole network was fine-tuned for 40 epochs with a lower learning rate of 0.00001. This much lower rate is a standard practice for fine-tuning, as it prevents the general-purpose features learned on ImageNet from being catastrophically forgotten during the update process (Yosinski et al. 2014). The patience values of five epochs for learning rate reduction and 20 for early stopping were chosen as common heuristics to balance efficient training time with allowing the model to fully converge.

Table 5.4. Architecture and computational characteristics of the four convolutional models trained to classify images of salmonids as 'healthy' or 'Saprolegnia spp.'. For each backbone we report the input resolution, the size of the added dense layer (Dense units), number of model layers unfrozen during the fine-tuning phase, the number of total and trainable parameters (at fine-tuning), and the approximate inference time per image measured on a P100 GPU.

Backbone	**Input size [px]**	**Dense units**	**Unfrozen layers to fine-tune**	**Total parameters**	**Trainable parameteres**	**Inference time per image [ms]**
ResNet50	224x224	128	98	23,850,242	15,216,002	104.35

Backbone	Input size [px]	Dense units	Unfrozen layers to fine-tune	Total parameters	Trainable parameteres	Inference time per image [ms]
MobileNetV3S	224x224	128	52	1,013,234	803,114	27.38
EfficientNetV2B0	224x224	256	198	6,247,762	5,707,446	105.05
EfficientNetV2S	300x300	512	363	20,988,258	18,876,042	412.38

We used k-fold cross-validation with k=5 to train each model architecture 5 times. The arithmetic mean and standard deviation for each evaluation metric were calculated using the implementations of mean and std in NumPy.

Model Evaluation

We assessed models using common metrics: precision (Eq. 5.1), recall (Eq. 5.2) and F1 (Eq. 5.3). We also calculated the Matthews Correlation Coefficient (MCC) (Eq. 5.4) (Chicco and Jurman 2020). These metrics denote relationships between the numbers of True Positives (TP), True Negatives (TN), False Positives (FP) and False Negatives (FN). True positives for the '*Saprolegnia* spp.' class represent the number of cases correctly identified as presenting with *Saprolegnia* spp. and true negatives the number of cases correctly identified as not presenting with *Saprolegnia* spp., as 'healthy'. Similarly, false positives, or type 1 errors, are images incorrectly classified as '*Saprolegnia* spp.' and false negatives, or type 2 errors, are the number of disease cases incorrectly classified as 'healthy'.

$$Precision = \frac{TP}{TP + FP} \tag{5.1}$$

$$Recall = \frac{TP}{TP + FN} \tag{5.2}$$

$$F1 = 2 \times \frac{precision \times recall}{precision + recall} \tag{5.3}$$

$$MCC = 2 \times \frac{TP \times TN - FP \times FN}{\sqrt{(\mathrm{TP} + \mathrm{FP})(\mathrm{TP} + \mathrm{FN})(\mathrm{TN} + \mathrm{FP})(\mathrm{TN} + \mathrm{FN})}} \tag{5.4}$$

Precision (Eq. 5.1) measures the proportion of correct positive predictions amongst all positive predictions made by the model. For example, if the model identifies 100 images as '*Saprolegnia* spp.', with 90 actually showing visible signs of *Saprolegnia* spp., the precision would be 90%. High precision indicates that when the model predicts a particular condition is present, it is usually correct. Recall (Eq. 5.2) measures the proportion of actual positive cases correctly identified by the model. For instance, if there are 100 images of fish with visible signs of *Saprolegnia* spp. in the dataset, and the model correctly identifies 80 of them, the recall would be 80%. High recall indicates that the model is successfully detecting most instances of the condition of interest. The F1 (Eq. 5.3) is the harmonic mean of precision and recall, balancing both measures. For example, in disease monitoring, we want to avoid both incorrectly identifying healthy fish as diseased (false positives, affecting precision) and missing cases of actual disease (false negatives, affecting recall). A high F1 indicates that the model maintains both good precision and good recall. The MCC (Eq. 5.4) produces a value between -1 and +1, representing the correlation between observed and predicted classifications. A coefficient of +1 represents a perfect prediction, 0 is no better than random guessing, and -1 indicates total disagreement. MCC calculates the correlation using all four categories of the confusion matrix (true positives, true negatives, false positives, and false negatives), making it robust for imbalanced datasets.

We focus on models with high recall for the '*Saprolegnia* spp.' class to minimise missed cases (false negatives), even if it might increase false positives. We also consider precision

for the disease class. To account for the class imbalance, we use the macro-average F1-score (the mean of both class-wise F1 scores) and MCC. From here on, we refer to class-wise metrics as $metric_{class}$, where 'metric' is one of the metrics, and 'class' is either 'healthy' or 'sapro' for *Saprolegnia* spp.

Cross-Evaluation Strategy

To enable fair comparisons between training strategies, we performed a cross-evaluation in which models trained on different tiered datasets were evaluated on identical held-out folds of two genus-specific evaluation sets (*Salmo* and *Oncorhynchus*). For each architecture and training strategy, we trained models using 5-fold stratified group cross-validation. For each fold, we evaluated the resulting model not only on its within-fold held-out data, but also on the corresponding held-out fold from the *Salmo* evaluation set and from the *Oncorhynchus* evaluation set. This produced fold-paired performance estimates on the same images, allowing direct comparison of training strategies when evaluated on a fixed target genus.

Comparison to a Random Model

To establish a simple, unbiased benchmark, we implemented a random baseline classifier that assigns each image to 'healthy' or '*Saprolegnia* spp.' with equal probability ($p = 0.5$). For each dataset (and for each cross-validation fold's held-out validation set), we generated random predictions for all images and computed the same evaluation metrics as for trained models ($precision_{sapro}$, $recall_{sapro}$, macro-average F1 and MCC). Because the random baseline is stochastic, we repeated this procedure 100 times. For each fold we then summarised the baseline as the mean metric value across repeats (and report the mean and 95% interval across repeats for descriptive purposes).

5.2.3 Qualitative and Quantitative Analysis

We used saliency methods to visualise which parts of an image most influenced a model classification decision and generated heatmaps that highlight those regions. We applied Grad-CAM, (Selvaraju et al. 2020) in TensorFlow, following the implementation in Chollet (2021) for the best performing model on all images in the *Salmo* genus-specific dataset. We categorised images as: correctly identified 'healthy' or '*Saprolegnia* spp.' images and

incorrectly identified 'healthy' or '*Saprolegnia* spp.' images. Through manual inspection of correctly (n=1208) and incorrectly (n=17) classified images and their corresponding Grad-CAM heatmaps, we investigated whether systematic patterns existed in the regions of the image that most influenced the model's classification decisions. To quantitatively investigate how image characteristics influenced model performance, we calculated several image quality and dimension metrics for each image. We then compared the distributions of these metrics between correct and incorrect classifications. Full methodological details and statistical analyses are provided in Supplementary Material, Chapter 5.

5.3 Results

5.3.1 Model Comparison

All performance comparisons below use the cross-evaluation in which models trained under different dataset strategies were evaluated on the same held-out folds of the Salmo and Oncorhynchus evaluation sets (Figures 5.3–5.4). The EfficientNetV2S architecture generally demonstrated the strongest performance across both evaluation sets, achieving the highest macro-average F1 and (in most cases) the highest MCC. EfficientNetV2B0 consistently ranked second, while the smaller MobileNetV3S and ResNet50 backbones generally produced lower metric scores.

On the Salmo evaluation set, the highest overall performance was achieved when models were trained on the genus-specific dataset ('*Salmo*, ≥10 photographs'). Using this training strategy, EfficientNetV2S attained a macro-average F1 of 0.930 ± 0.026 and an MCC of 0.859 ± 0.051 (Figure 5.3b-c). It also achieved a $recall_{sapro}$ of 0.918 ± 0.038 (Figure 3.5a) and a $precision_{sapro}$ of 0.862 ± 0.056. This exceeded the performance of the same architecture trained on the broader 'All photographs' dataset (macro-average F1 0.910 ± 0.024; MCC 0.826 ± 0.054). However, the combined training strategy ('Taxa with ≥10 photographs in both classes') proved a competitive alternative, with EfficientNetV2S achieving the highest $precision_{sapro}$ (0.908 ± 0.055) of all models, while maintaining comparable performance across macro-average F1 (0.922 ± 0.036) and MCC (0.845 ± 0.075). Across all architectures and training strategies, performance on the Salmo evaluation set exceeded the random baseline across $recall_{sapro}$, macro-average F1 and MCC (Figure 5.3).

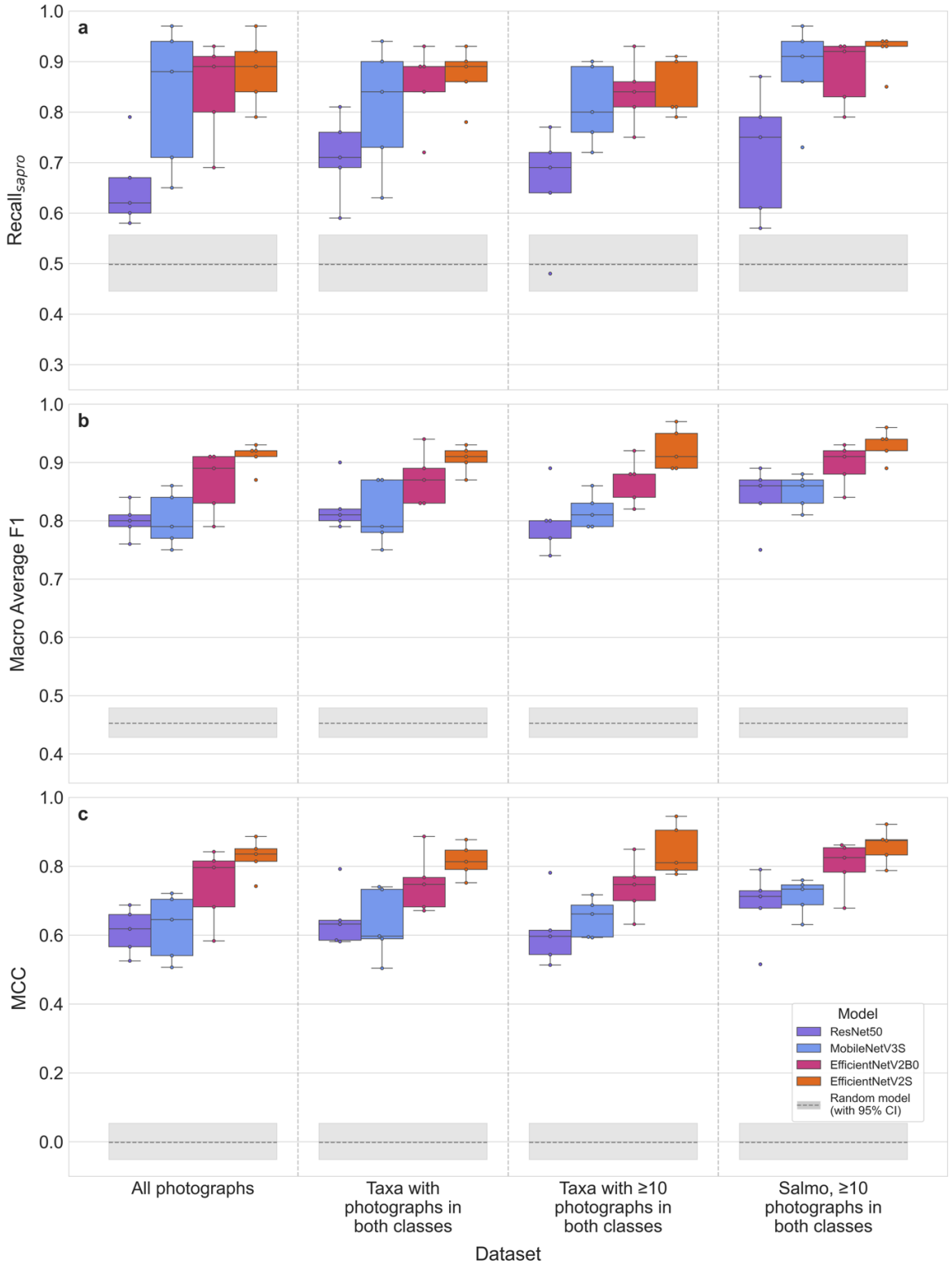


Figure 5.3. Performance metrics on the *Salmo* data. Comparison of recall$_{sapro}$, macro-average F1 and MCC for different models trained on the different tiered datasets and tested on the *Salmo* genus-specific data. A random model with 95% confidence level, in grey, for each metric and dataset was calculated by running a model that randomly predicts the 'healthy'

or '*Saprolegnia* spp.' class with equal probability 100 times. Box plots show median, interquartile range, whiskers indicate 1.5 x IQR, with individual points for each model fold run overlaid.

Performance was lower on the *Oncorhynchus* evaluation set, with greater variation between training strategies (Figure 5.4). For EfficientNetV2S, genus-specific training ('*Oncorhynchus*, ≥10 photographs') yielded the highest $precision_{sapro}$ (0.478 ± 0.073) and macro-average F1 (0.734 ± 0.044). In contrast, training on the broader 'All photographs' dataset increased sensitivity, producing the highest $recall_{sapro}$ (0.674 ± 0.073), compared with 0.612 ± 0.072 for the genus-specific model. Regarding the random baseline, while $recall_{sapro}$ for the ResNet50 architecture overlapped with the random confidence interval across all training strategies (Figure 5.4a), the MCC values for all trained models remained consistently positive and above the random baseline band (Figure 5.4c).

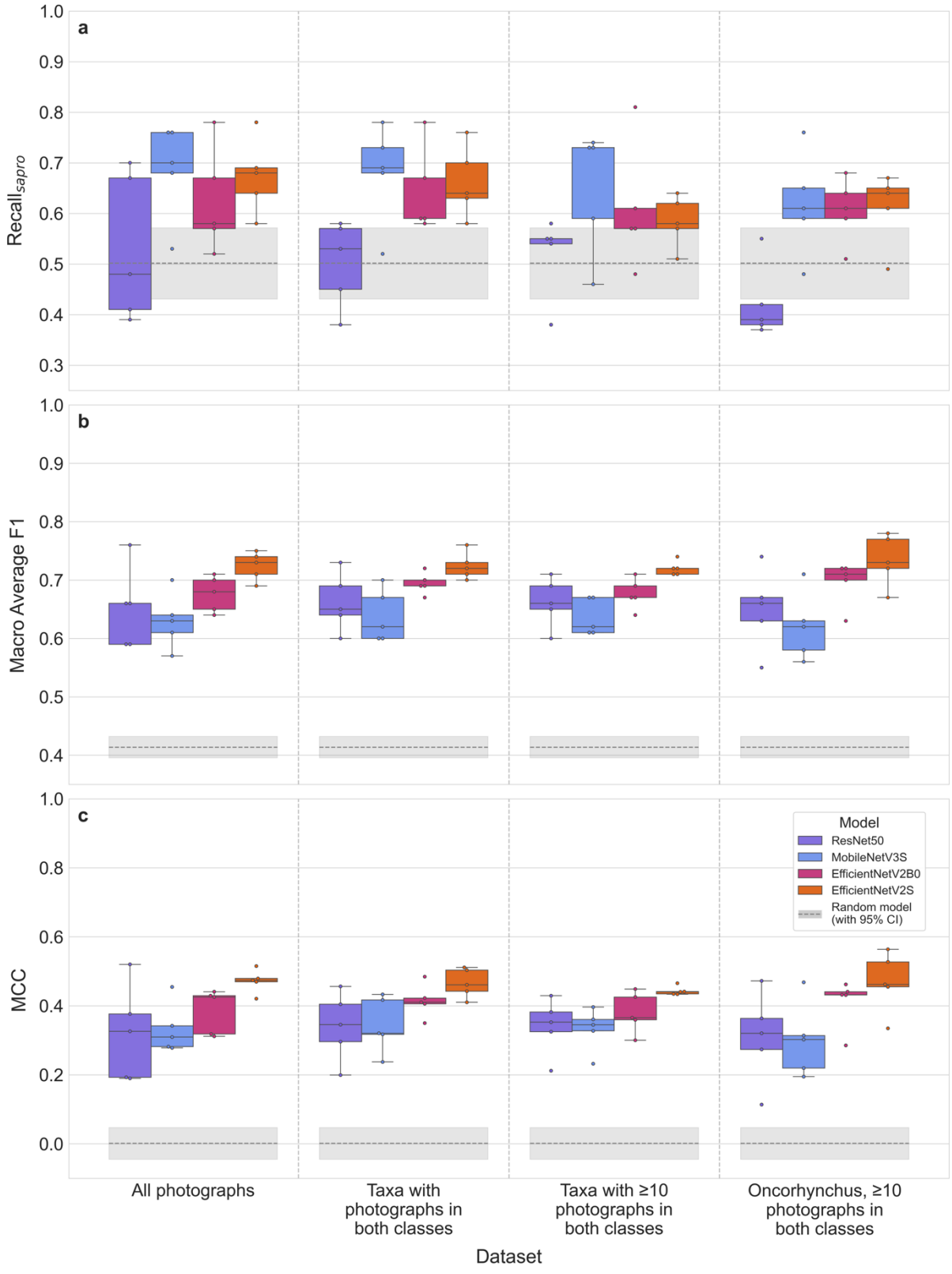


Figure 5.4. Performance metrics on the *Oncorhynchus* data. Comparison of recall$_{sapro}$, macro-average F1 and MCC for different models trained on the different tiered datasets and tested on the *Oncorhynchus* genus-specific data. A random model with 95% confidence level, in

grey, for each metric and dataset was calculated by running a model that randomly predicts the 'healthy' or '*Saprolegnia* spp.' class with equal probability 100 times. Box plots show median, interquartile range, whiskers indicate 1.5 x IQR, with individual points for each model fold run overlaid.

5.3.2 Qualitative and Quantitative Analysis

Grad-CAM analysis revealed that surface reflections were consistently responsible for misclassifications for EfficientNetV2S, particularly when surface reflections obscured fish features (Figure 5.5). Manual inspection highlights that the model can correctly focus on infection, with the strongest activation (shown in red to turquoise) around the dorsal fin and midsection where the infection was visible (Figure 5.5e). Similarly, when correctly classifying a healthy brown trout the model appropriately concentrated on the fish's body, with the highest activation along the main body and adjacent areas (Figure 5.5f). However, where water surface reflections created both bubbles above and reflective patterns below the fish (Figure 5.5c-d) the corresponding heatmap (Figure 5.5g-h) reveals that the model focused primarily on these water disturbances leading to misclassifying these healthy fish as diseased.

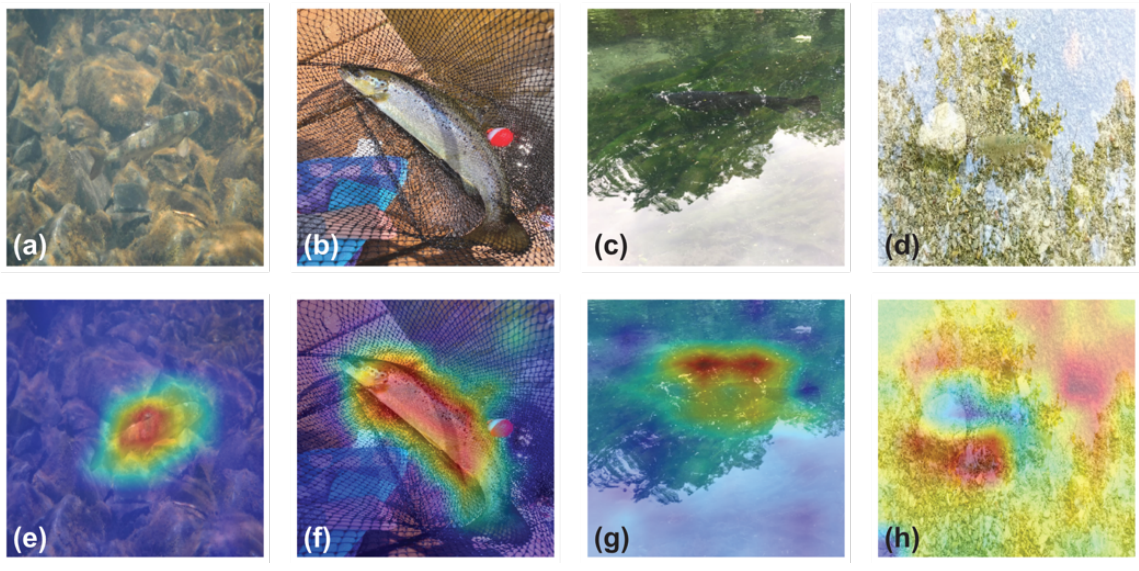


Figure 5.5. Grad-CAM heatmaps. Comparison of a subset of images used for EfficientNetV2S classification on the 'Salmo, ≥ 10 photographs in both classes' dataset (a-d) and their

corresponding Grad-CAM heatmap overlays (e-h). (a) Atlantic salmon (*Salmo salar*) correctly classified as infected with '*Saprolegnia* spp.', (b) Atlantic salmon (*Salmo salar*) correctly classified as 'healthy', (c and d) Brown trout (*Salmo trutta*) incorrectly classified as '*Saprolegnia* spp.', and corresponding Grad-CAM heatmap overlays (e-h). See Supplementary Material, Chapter 5 for photograph attribution.

The quantitative analysis of image characteristics found no statistically significant association between classification outcome and global image quality metrics. Comparing correctly and incorrectly classified images using a Mann-Whitney U test on the full dataset revealed no significant differences in sharpness ($p=0.366$), perceptual quality (BRISQUE, $p=0.266$; NIQE, $p=0.787$), or image dimensions ($p>0.9$). Full details of this analysis are provided in Supplementary Material, Chapter 5.

5.4 Discussion

Our results demonstrate the potential for computer vision to support disease surveillance in wild fisheries. By performing a rigorous cross-evaluation on held-out folds of the *Salmo/Oncorhynchus* evaluation sets, we found that deep learning models outperformed random classification on macro-average F1 and MCC across both evaluation sets (Figures 5.3-5.4), with the strongest performance on the *Salmo* genus-specific dataset. Our work demonstrates the potential for rapid and extensive surveillance, mindful of potential methodology pitfalls, with classification potentially improved with more training images.

Previous work has reported near-perfect classification of fish diseases, including *Saprolegnia* spp. (models: VGG16, MobileNetV2 and InceptionV3 (Biswas et al. 2024). However, manual inspection of the images used (https://www.kaggle.com/datasets/subirbiswas19/freshwater-fish-disease-aquaculture-in-south-asia), revealed potential overfitting due to augmented versions of images from the training set constituting those in the test set. Kumaar et al. (2024) also achieved high performance (models: InceptionV3, VGG16 and a custom FishNetCNN) on the same dataset, expanded with additional images, but had inconsistent sample sizes and possible augmented training samples in the test set. Although our best model (EfficientNetV2S) achieved lower metrics for classification of *Saprolegnia* spp. than these works (Biswas et al.

2024; Kumaar et al. 2024), we adhered to stricter validation protocols, reducing bias and improving real-world application for monitoring diseases in the field.

5.4.1 Model Performance and Technical Considerations

The strong performance of EfficientNetV2S can most likely be attributed to its ability to process higher resolution images (300x300 pixels versus 224x224 pixels) enabling detection of subtle disease features. A key finding from our cross-evaluation was the influence of training data composition. For the *Salmo* dataset, which benefitted from a more favourable class balance (highest ratio of disease to healthy images), the specialist model performed best, refuting the hypothesis that a generalist model trained on all available data would necessarily outperform a specialist model due to volume alone. This suggests that when a target taxon is well-represented, adding images from other genera introduces additional visual variation that may degrade performance. However, the combined training strategy (taxa with ≥10 photographs) proved highly competitive, achieving the highest precision, which suggests that curated multi-taxa datasets can improve model robustness against false positives. Conversely, for the *Oncorhynchus* dataset, which had fewer disease examples, the generalist 'All photographs' model improved recall for the '*Saprolegnia* spp.' class by leveraging features learned from the *Salmo* examples. This highlights a strategic trade-off for surveillance: specialist models are preferable where data is more balanced to maximise MCC and precision, while generalist models can boost sensitivity for data-poor taxa via transfer learning, albeit with a trade-off in precision.

Grad-CAM visualisations suggested the model focused on relevant anatomical features in correctly classified images, but when misclassifying did so most likely so due to water surface reflections (Figure 5.5), thereby providing guidance for future work that model performance, i.e. disease classification, could be improved by restricting the type of images used.

All models, particularly MobileNetV3S, overpredicted '*Saprolegnia* spp.' leading to higher $recall_{sapro}$ (proportion of infected fish images that are correctly identified) but lower $precision_{sapro}$ (proportion of correctly identified infected fish among all fish identified as being infected). Oversampling to address class imbalance changes the class distribution in the training data and this can drive the models to overpredict the minority class. Although

we mitigated this using image augmentation to increase the size of our training datasets, alternative strategies like multi-branch networks, as suggested for visual recognition of animal species in camera-trap images (Cunha et al. 2023), could enhance performance.

An important consideration for practical application is the trade-off between accuracy and computational efficiency, as well as the intended use case. This tool could be deployed in two main ways. First, for post-hoc analysis, automatically screening large image collections from online repositories to flag potential cases for follow-up by managers. Where geotags and timestamps are available, such screening could support spatiotemporal exploration of apparent disease signals, although any inferred patterns would require careful adjustment for uneven observation effort and reporting biases. Second, as an *in situ* mobile application to support rapid field triage. Our model comparison was designed in part to explore such deployment trade-offs: lightweight architectures such as MobileNetV3S offer fast inference, while EfficientNetV2 variants provide improved performance. The inference times reported here (Table 5.4) were measured on an Nvidia P100 GPU and therefore indicate relative computational cost rather than expected mobile latency; real-world on-device performance would require conversion to a mobile runtime (e.g., TensorFlow Lite/Core ML) and device-specific benchmarking and optimisation (e.g., quantisation). In both deployment modes, predictions should be treated as a screening signal rather than diagnosis, with confirmatory diagnostics (e.g., molecular testing) required for case confirmation, and decision thresholds selected to reflect the relative costs of false negatives and false positives.

5.4.2 Data Challenges in Disease Detection

Developing robust disease detection models is hindered by the difficulty of obtaining expert annotations for images with confirmed disease. Citizen science data introduces noise and geographic bias (Edwards et al. 2021) and in the current work, UK-based images were overrepresented. Many images were shared by anglers, and although these stakeholders offer an opportunity to acquire many images, they may be less likely to take 'trophy' images of diseased fish or only share pictures of fish in advanced stages of disease, so creating bias. Combined with inconsistent taxonomy across sources and variable image quality, the complexities of building representative datasets are apparent using internet harvested images. Computer vision techniques such as augmentation (making random changes to

existing images to increase dataset size) increase sample size, but risk inadvertently amplifying existing biases (Shorten and Khoshgoftaar 2019). For our study, an important next step to improve model performance and generalisability could be to expand the labelled dataset by annotating a larger portion of the thousands of unlabelled salmonid images acquired during our initial data collection. With *post-hoc* image processing not being a complete solution, engaging and training stakeholders (here, anglers) to submit images of both healthy and unhealthy-looking fish could help overcome some biases and class imbalances. Similar citizen science approaches have worked well for surveillance of other wildlife diseases, such as sarcoptic mange in foxes, *Vulpes vulpes* (Scott et al. 2020), although they did not use computer vision. Open access image repositories (iNaturalist, Flickr, GBIF), offer great opportunities to develop structured citizen science programmes with standardised imaging protocols (August et al. 2020b). Indeed, iNaturalist and Flickr were valuable resources for collecting a large dataset of 4,526 salmonid images for our study.

Online platforms offer great potential to collate a large number of images, providing a cost-effective alternative to traditional field surveys; they capture valuable metadata, including date, time, location, which is crucial for disease surveillance and 'research grade' images have been taxonomically identified. An additional benefit of using online repositories is that users have explicitly agreed to share their observations under Creative Commons licenses, aligning with best practices including Findable, Accessible, Interoperable, and Reusable (FAIR) data (Wilkinson et al. 2016). However, leveraging citizen science data still requires careful curation and quality control due to the inherent biases in these data (Brown and Williams 2019).

Academic or practitioner curated databases of confirmed disease cases, such as those used in our study, offer a potential solution for training models. However, these databases are often not open access. While there is a growing infrastructure for sharing images (e.g., Kaggle, Zenodo) and increasing calls for collaboration in building species-specific disease databases (Nunes et al. 2020), challenges remain in transforming collated data into resources that effectively meet researchers' needs. Good annotation practices and standardised protocols are needed to make these datasets broadly useful. For example, 'SalmonScan' (Ahmed 2024), although a large dataset (1,208), constitutes augmented

images from 24 uninfected and 91 infected fish, and lacks details about species identification and infection types.

5.4.3 Challenges in Wild Fish Disease Detection

Specific to disease detection in wild fish, the Grad-CAM analysis revealed misclassification of fish underwater due to water surface features. Incorporating fish detection using tools like 'megafishdetector' (Yang et al. 2023) could help isolate individual specimens in complex images. Pre-processing techniques could reduce reflection effects crucial for minimally invasive *in situ* monitoring. Furthermore, given the differing performance on Salmo versus Oncorhynchus evaluation sets, a hierarchical model incorporating taxonomic data (Elhamod et al. 2022), may improve classification accuracy by explicitly modelling the visual differences between genera.

5.4.4 Implications for Disease Surveillance

Our results demonstrate the potential for computer vision to transform disease surveillance in wild fish populations. While our models cannot replace traditional diagnostic methods, as confirmation of *Saprolegnia* spp. infection requires a confirmed molecular diagnosis (van West and Beakes 2014), they offer a valuable tool for rapid, large-scale screening. This approach could help identify potential disease outbreaks earlier, enabling more targeted application of confirmatory tests.

The success with *Saprolegnia* spp. suggests potential applications for other visually distinctive diseases not only in fish but also in other animals, such as mange in *Vulpes vulpes* (Scott et al. 2020). Integration with spatiotemporal metadata, available for all research grade iNaturalist observations, could provide insights into disease dynamics and environmental drivers of outbreaks if large enough datasets could be acquired. However, such datasets would likely contain inherent biases, as observation frequency often correlates with human population density and accessibility of sites (Geurts et al. 2023). Additionally, temporal biases may arise from seasonal variations in observer effort and species visibility. These sampling biases would need careful consideration when interpreting any apparent patterns in disease occurrence or distribution.

This work represents a significant step toward automated disease surveillance in wild fish populations, demonstrating both the potential and challenges of computer vision approaches. Overall, this work highlights the transformative potential of computer vision for disease surveillance in fish, but also other visually distinct wildlife diseases, while also underscoring the need for continued refinement and careful integration with existing methods.

CHAPTER 6 :

Deep Learning Model Confidence as a Proxy for Disease-Severity: a Case Study of Rumpwear in Possums

Agnethe Seim Olsen[1], Sarah E Perkins[1], Jo Cable[1], Elise M. Ringwaldt[2], Jessie Buettel[2], Barry W. Brook[2]

[1]School of Biosciences, Cardiff University, Cardiff, CF10 3AX, UK

[2]School of Natural Sciences, University of Tasmania, Hobart, Australia, Private Bag 55, Hobart TAS 7001

Abstract

Traditional wildlife disease surveillance is often resource-intensive, and computer vision offers a promising avenue for more efficient, automated monitoring. Here, we evaluate deep learning models for classifying 'rumpwear', an externally visible condition in common brushtail possums (*Trichosurus vulpecula*), using camera trap images. Rumpwear ranges from hair breakage to complete hair loss, and its visually gradable nature makes it an ideal case study for moving beyond simple disease presence/absence detection. Our best-performing supervised model achieved 95.2% classification accuracy. Importantly, we demonstrate that the model's output probability for the 'Disease' class correlates significantly with expert-assigned severity scores on held-out test images, validating its use as a proxy for a disease-severity index. Gradient-weighted Class Activation Mapping showed the model concentrated on the rump, which was the relevant region for rumpwear clinical signs. Developing deep learning models that can rapidly classify and quantify disease-severity from images is a useful tool for enhancing existing disease monitoring projects and offers vital infrastructure for future programmes.

6.1 Introduction

Effective disease surveillance is crucial globally. Yet achieving comprehensive monitoring faces significant challenges, particularly for wildlife populations where efforts often rely on opportunistic reporting rather than systematic surveys (Phelps et al. 2019; Watsa 2020). While frameworks like the World Organisation for Animal Health (WOAH) collate reports on notifiable diseases, primarily focusing on livestock and major epizootics through methods like clinical observations and sentinel monitoring, obtaining comparable data for many wild species remains difficult (Stallknecht 2007; Watsa 2020). Consequently, understanding

disease dynamics in wildlife often depends on passive detection, such as reports of dead or sick animals from the public, for example hunters, amateur naturalists or anglers.

Deep learning models have proven effective for image-based disease detection and classification in agriculture, particularly plant pathology (Francis and Deisy 2019; Li et al. 2020), and medical diagnostics (Esteva et al. 2021; Rana and Bhushan 2023). Despite this success, and the rapid growth of camera-trap image datasets (Swanson et al. 2015), few studies apply these techniques to wildlife disease surveillance (Christin et al. 2019; Jarić et al. 2020; Poulin et al. 2021). Many animal diseases manifest externally visible signs, presenting a clear opportunity to leverage computer vision (CV) for automated detection from photographs or videos. Encouragingly, specific applications are emerging: for instance, CV has been used to classify skin lesions in bottlenose dolphins (*Tursiops truncatus)* from photographic surveys (Murphy et al. 2025) and to detect Devil Facial Tumour Disease (DFTD) in Tasmanian devils (*Sarcophilus harrisii*) from camera trap images (Nurçin et al. 2024). Recently, Ringwaldt et al. (2025) also used a CNN to identify rumpwear in common brushtail possums for broad-scale distribution mapping. Furthermore, the potential utility of image repositories for assessing disease in aquatic species, such as salmonids, has found that models can distinguish healthy from diseased animals (Olsen et al. 2026, Chapter 5). Despite these advancements, broad application and methodological refinement across diverse wildlife-disease systems are needed to fully realise the potential of image-based surveillance.

One condition presenting with visually apparent signs in Australian marsupials is rumpwear, also known as ‘rumpy possum’ or ‘rumpiness’. It affects common brushtail possums (*Trichosurus vulpecula*) across Australia, including Tasmania (Ringwaldt et al. 2025). Rumpwear has also been described in mountain brushtail possums (*Trichosurus cunninghami*, see Hufschmid et al. 2010) and common ringtail possums (*Pseudocherius peregrinus*, see Ringwaldt et al. 2022). Rumpwear manifests as bilateral hair damage on the lumbosacral region, ranging from changes in fur colouration due to hair breakage to, in rare cases, complete hair loss (Hufschmid et al. 2010; Ringwaldt et al. 2022). While the precise aetiology remains unclear, it is hypothesised to involve hypersensitivity to an irritant, possibly ectoparasites, leading to mechanical hair damage from overgrooming (Hufschmid et al. 2010). The visually gradable nature of these signs, along with evidence suggesting links

to host factors, seasonality, and population density (Hufschmid et al. 2010; Ringwaldt et al. 2025), makes rumpwear a suitable model syndrome for developing and evaluating computer vision tools beyond simple presence/absence detection, particularly for assessing varying levels of affliction.

Building on the epidemiological findings and initial image classification efforts of Ringwaldt et al. (2025), this study used a dataset of 7780 images of common brushtail possums from the same Tasmanian camera trap network (Vaughan et al. 2022; Brook et al. 2025; Ringwaldt et al. 2025). These images were expertly scored for the presence and apparent severity ('Mild signs' versus 'Obvious signs') of rumpwear by Ringwaldt and Brook, as detailed in Ringwaldt et al. (2025). While Ringwaldt et al. (2025) developed a CNN for broad-scale rumpwear detection and distribution mapping, our study provides a novel methodological framework for moving beyond simple presence/absence classification. Specifically, this paper aims to: 1) comprehensively evaluate different deep learning model configurations (including pre-training and fine-tuning strategies) to accurately classify rumpwear presence, comparing model performance to expert labels; 2) investigate whether model outputs (e.g., prediction probabilities) can serve as a robust proxy for a nuanced, expert-scored disease-severity index, aligning with the original multi-category expert scores; and 3) explore the utility of semi-supervised learning techniques to potentially reduce the reliance on extensive manual labelling for training effective classifiers in such wildlife-health studies.

6.2 Materials and methods

6.2.1 Data Acquisition and Preparation

We sourced possum images from the University of Tasmania's state-wide camera-trap network. This network and its operational details (914 unique camera sites, active 2016-2021) are described in Vaughan et al. (2022) and Ringwaldt et al. (2025). Most units employed white-flash illumination, producing colour night images that allowed clearer visual grading of rumpwear in this nocturnal species. Initial image processing, including animal detection using MegaDetector (Beery et al. 2019) and subsequent species classification to identify common brushtail possums was done using the "Mega-Efficient Wildlife Classifier" (MEWC), as described by Brook et al. (2025).

Two experienced observers (co-authors Ringwaldt and Brook) annotated 7779 images for signs of rumpwear. Each image was assigned to one of four initial categories based on visual evidence: 1) Healthy: no signs of rumpwear; 2) Mild signs: potential rumpwear indicated by hair breakage or light grey/cream fur on the rump; 3) Obvious signs: clear breakage or loss of fur on the rump; and 4) Occluded: rump not visible or image quality insufficient for assessment (Figure 6.1).

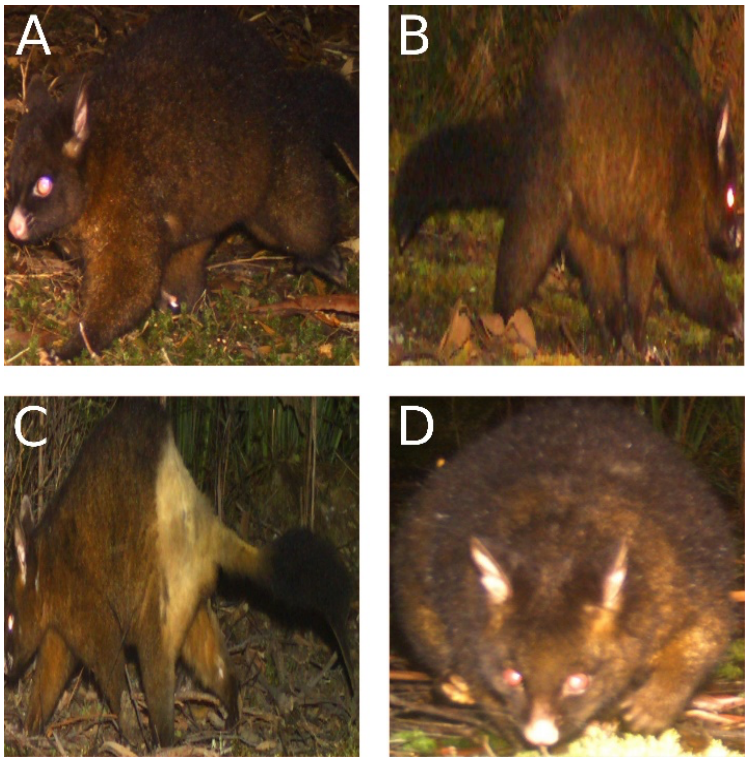


Figure 6.1. Camera trap images of brushtail possums exemplars, corresponding to the classes defined in this study. (A) Healthy (B) mild signs of rumpwear (C) obvious signs of rumpwear, and (D) possum with rump occluded.

For the primary model training, the initial expert-annotated categories of 'Mild signs' and 'Obvious signs' were consolidated into a single 'Disease' class. This was done to ensure a sufficient number of training examples for the overall diseased category, given the potential for fewer samples in more granular severity sub-classes, and to account for the inherent subjectivity and continuous nature of visual severity assessment. The original 'Mild' and 'Obvious' distinctions were, however, retained for a subsequent *post-hoc* analysis to investigate whether model outputs could serve as a proxy for disease-severity (see Section 6.2.8).

6.2.2 Dataset Partitioning and De-duplication

To ensure the integrity and independence of the dataset, we applied a multi-step partitioning and de-duplication pipeline. For each class, image embeddings were first computed for all images using the clip-ViT-B-32 model from OpenAI (Radford et al. 2021) via the sentence-transformers library (Reimers and Gurevych 2019). We computed pairwise cosine similarity within each class to flag duplicates.

Pairs with cosine similarity ≥ 0.999 were treated as exact duplicates. Within each group of identified exact duplicates (caused by overlapping object detections by MegaDetector), all but one image were excluded from the final dataset, resulting in 7657 unique images after duplicate removal. Near-duplicate sequences (scoring 0.97 – 0.999) likely represented time series of the same individual animal captured in quick succession. We placed all near-duplicate sequences solely in the training set to prevent images from the same animal encounter appearing in both the training and evaluation (validation or test) sets, thereby reducing the risk of data leakage.

Following the removal of exact duplicates and the allocation of near-duplicate sequences to the training set, the remaining unique images within each class were statically partitioned. The target split was 80% for training, 10% for validation, and 10% for testing, based on the post-duplicate-removal image count for each class. This procedure resulted in the final dataset splits shown in Table 6.1.

Table 6.1. The number of images in the training, validation and test set for each of the classes ‘Healthy’, ‘Disease’ and ‘Occluded’

Class	**Training**	**Validation**	**Test**	**Total**	**Proportion (%)**
Healthy (no rumpwear)	3177	397	397	3971	51.9
Disease (signs of possible or obvious rumpwear)	1329	166	167	1662	21.7
Occluded (rump of animal occluded)	1619	202	203	2024	26.4

Class	Training	Validation	Test	Total	Proportion (%)
Total	6125	765	767	7657	100.0

6.2.3 Model Architecture and Pre-training Strategies

We chose EfficientNetV2S convolutional neural network (CNN) architecture (Tan and Le 2021) for its high accuracy to computational efficiency ratio in image recognition tasks. All models were implemented using Keras 3 with a JAX backend in Python 3.10.13. A transfer learning approach was adopted, with model weights initialised from two distinct pre-training sources to enhance feature extraction.

First, we used ImageNet pre-training for generic visual features (Deng et al. 2009). The second source involved custom wildlife pre-trained weights, generated to leverage domain-specific features relevant to camera trap imagery. These custom weights were produced by adapting the workflow from the MEWC pipeline (Brook et al. 2025), with the associated codebase available on GitHub (http://github.com/zaandahl/mewc). The MEWC framework is designed for customised wildlife image classification using deep learning. Brook et al. (2025) demonstrated the MEWC workflow's capability by training models (including EfficientNetV2S) on a dataset of 10 common Tasmanian wildlife species/classes, reportedly achieving high classification accuracies (e.g., their EfficientNetV2S model achieved 99.48% accuracy and a test loss of 0.0023 on this multi-species task).

To prevent leakage, we first fine-tuned MEWC on nine non-possum Tasmanian mammal species. The resulting weights from this nine-species classification model then served as our custom wildlife pre-trained weights, providing CNN features adapted to the specific camera network and Tasmanian environmental context.

For all model configurations, we passed the features extracted by the pre-trained EfficientNetV2-S base model to a custom classification head to map the learned feature representations to the specific target classes. This comprised of a GlobalAveragePooling2D layer to summarise the spatial feature maps, followed by a Dropout layer. Dropout layers were included to reduce overfitting by randomly deactivating a subset of neurons during training, encouraging the model to learn more generalisable feature representations. The

subsequent layers included a Dense layer with 256 units, an Exponential Linear Unit (ELU) activation layer, and then a second Dropout layer. The head terminated with a final Dense classification layer with softmax activation, which predicted probabilities for the 'Healthy', 'Disease', or 'Occluded' categories.

6.2.4 Supervised Training Protocol

We investigated four primary supervised training configurations, varying both the source of pre-trained weights (ImageNet or custom wildlife) and the initial training strategy (frozen base then fine-tune, or fine-tune only). This comparison allowed us to determine whether the model benefits more from domain-specific prior knowledge (local wildlife) than generic features (ImageNet), and to assess whether the entire network needs to be updated to capture the fine-grained visual characteristics of rumpwear.

For the "frozen then fine-tuned" (fr_ft) configurations, the model was initialised with either ImageNet or custom wildlife pre-trained weights. During the initial frozen training phase, the weights of the base CNN were kept fixed, and only the custom classification head was trained for 10 epochs. Following this, the top block of the EfficientNetV2S base was unfrozen, and the entire model was fine-tuned for a further 28 epochs, divided into four progressive stages of seven epochs each.

In the "fine-tuned only" (ft) configurations, the model was initialised with the respective pre-trained weights, and the top block of the base model was unfrozen immediately. The model was then fine-tuned for 28 epochs, also using progressive training stages.

All training runs used the AdamW optimiser with a learning rate of 1e-4 and weight regularisation (1e-4), together with an exponential learning rate scheduler. To address class imbalance, the training data was balanced by resampling (without replacement) 1300 images per class for each epoch. Input images were resized to 384 × 384 pixels, and data augmentation was applied using RandAugment (Cubuk et al. 2020) with three augmentations per image and progressive augmentation magnitudes across the four fine-tuning stages (0.2, 0.4, 0.6, and 0.8, respectively). To further regularise the model during the fine-tuning stages (mitigating overfitting), the rate for the top dropout layer (located after the ELU activation in the classification head) was progressively increased across the four

stages (0.1, 0.2, 0.3, and 0.4, respectively). Training was performed with a batch size of 16. Model checkpoints were saved based on the lowest validation loss observed during training.

6.2.5 Model Evaluation Metrics

We assessed performance with the standard quartet of metrics: accuracy, precision, recall, and F1-score (definitions in Géron 2022, Chapter 3). Accuracy can be a misleading metric for this task because the dataset is imbalanced; for instance, a model that only ever predicts 'Healthy' would still achieve 90% accuracy if just 10% of possums were diseased. In disease surveillance, the cost of failing to identify a diseased animal (a false negative) is particularly high. High recall for the 'Disease' class means the model successfully identifies most true disease cases, generating few false negatives. As such, we treat accuracy as ancillary and base our conclusions on F1 and recall.

6.2.6 Model Selection for Subsequent Analysis

To ensure a fair and unbiased evaluation, we selected a single supervised model for all subsequent analyses, including the semi-supervised learning protocol and the disease-severity investigation. Model selection was based on a comparative evaluation of the four trained models on a held-out validation set, a standard practice to prevent information leakage from the test set during model development (Chollet 2021). We considered both discrimination performance (including precision for the 'Disease' class) and overall calibration, as reflected in the validation loss, because a key aim of the study was to use the model's continuous output probability as a proxy for disease severity. In particular, we prioritised models with lower validation loss, as these are more likely to produce reliable and interpretable probability scores across the severity spectrum, which is essential for generating high-quality pseudo-labels and for analyses relying on model output probabilities (Lin et al. 2020; Sohn et al. 2020).

6.2.7 Semi-Supervised Learning via Pseudo-Labelling

The scarcity of high-quality labelled data in ecological image analysis motivates the use of semi-supervised and self-supervised learning approaches (van Engelen and Hoos 2020). To investigate if model performance could be improved with unlabelled data, we adopted a semi-supervised learning (SSL). Among the various semi-supervised learning strategies, wrapper methods such as self-training (also known as iterative pseudo-labelling) are well-

established and widely used due to their simplicity and flexibility. These methods allow unlabelled data to be incorporated into model training with minimal modification to existing supervised algorithms, making them especially attractive for practical applications where ease of implementation and interpretability are important considerations (van Engelen and Hoos 2020). While more complex semi-supervised learning methods exist, there is no single best approach for all problems, and empirical evaluation is essential to determine the most suitable method for a given research question. In this study, we adopted an iterative pseudo-labelling strategy, balancing methodological rigour with practical feasibility, to investigate its utility in training effective rumpwear classifiers with reduced reliance on extensive manual labelling.

The iterative pseudo-labelling protocol was done with the elected supervised model over three iterations. In each iteration, the current teacher model generated pseudo-labels for the unlabelled dataset. Only images where the model's prediction confidence for a specific class surpassed a given threshold were assigned pseudo-labels. This confidence threshold was set to 0.99, 0.85, and 0.70 for the three successive iterations, gradually allowing more, potentially less certain, pseudo-labels to be incorporated.

Following pseudo-label generation in each of the three iterations, the teacher model was retrained using curriculum learning strategy (Bengio et al. 2009) where a model learns from easier examples before proceeding to more difficult ones. In our implementation, prediction confidence served as a proxy for example difficulty. Within each iteration, pseudo-labels meeting that iteration's confidence threshold were sorted by confidence. Training then proceeded through four curriculum stages: the expert-labelled data was first combined with the top 25% of these pseudo-labels, then progressively with the top 50%, 75%, and finally all 100%. Each of these four curriculum-stage datasets was used to fine-tune the model for 10 epochs, employing the same progressive training settings (dropout, augmentation magnitudes) as the supervised models. The model resulting from the final curriculum stage of an iteration became the teacher for the next.

6.2.8 Assessing Model Output as a Proxy for Disease-Severity

To investigate whether the outputs from the trained deep learning model could serve as an automated proxy for expert-assigned disease-severity scores (categorised as 'Mild signs' and 'Obvious signs'), a series of analyses were conducted. For severity, images with a ground truth label of 'Occluded' were excluded; this was because the rump is not visible in such images, meaning severity cannot be assessed by an expert, and therefore cannot be assessed for health status.

First, to understand general model calibration, we examined the distribution of the model's predicted softmax probability for its chosen class, grouped by the ground truth label ('Healthy', 'Disease', 'Occluded') across all data splits. A similar confidence analysis was also done on the large, unlabelled dataset by applying the final trained model to generate predictions, to understand the model's output distribution on a dataset representative of a real-world monitoring scenario.

To assess the relationship between model output and disease-severity more directly, several visualisations focused on the model's softmax probability for the 'Disease' class (P(Disease)). Violin plots were generated to compare the distribution of P(Disease) across test set images with ground truth labels of 'Healthy', 'Mild signs', and 'Obvious signs'. To statistically test for differences in P(Disease) between these three groups, we used a Kruskal-Wallis H test, a non-parametric approach suitable for comparing multiple groups without assuming a normal distribution of the data. Post-hoc analysis was subsequently performed using Dunn's test with Bonferroni correction to identify which specific group pairs were significantly different. Complementing this, a 2D scatter plot visualised P(Healthy) against P(Disease) for each non-occluded test set image, with points coloured by their ground truth label ('Healthy', 'Mild signs', 'Obvious signs') to identify potential separation in this two-dimensional probability space.

Finally, to explore whether the model's learned internal representations inherently captured severity-related information, dimensionality reduction techniques, specifically Uniform Manifold Approximation and Projection (UMAP), t-distributed Stochastic Neighbor Embedding (t-SNE; van der Maaten and Hinton 2008), and Principal Component Analysis (PCA), were used. These techniques were applied to two distinct sets of features from the non-occluded test set images: first, to the three-

dimensional vector of softmax output probabilities ([P(Healthy), P(Disease), P(Occluded)]), and second, to the 256-dimensional embeddings from the model's 'compression_bottleneck' layer. The resulting two-dimensional plots were coloured by the ground truth label ('Healthy', 'Mild signs', 'Obvious signs') to assess whether these categories formed distinct clusters, which would indicate that the model had learned separable features related to disease-severity. To quantitatively evaluate this clustering, we calculated the silhouette coefficient (Rousseeuw 1987). The silhouette score measures how similar a data point is to its own cluster compared to other clusters, where a higher score indicates that clusters are dense and well-separated. To further validate the reliability of the model's 'Disease' class probability, we also performed a calibration analysis (Guo et al. 2017). For this, we evaluated the model's performance on the binary task of distinguishing 'Healthy' from ‘Disease’ cases in the non-occluded test set. We calculated the Brier score loss (Brier 1950) and Expected Calibration Error (ECE), and generated a combined plot showing both the calibration curve and a histogram of the prediction confidences.

We analysed the Gradient-weighted Class Activation Mapping (Grad-CAM; Selvaraju et al. 2020) heatmaps for all non-occluded test images (n=564) to provide a quantitative assessment of the model's visual focus. Each heatmap was visually classified as either 'cohesive', where activation was clearly concentrated on the possum's rump, or 'fragmented', where activation was diffuse, scattered, or focused on other image elements. We then calculated the proportion of cohesive and fragmented heatmaps across different groups, such as correct versus divergent classifications and different levels of model confidence.

6.2.9 Model Interpretability: Identifying Key Image Regions

To understand the visual basis for the classification decisions made by the wildlife_ft model, we used Grad-CAM. Grad-CAM is a visualisation technique that produces a "heatmap" highlighting the specific regions in an input image that were most influential in the model's prediction for a given class (e.g., 'Disease').

Rather than using a simple random sample, we implemented a targeted, scenario-based strategy to select the most informative images from the test set. This approach allowed us

to probe the model's behaviour under various conditions. We generated Grad-CAM heatmaps for the model's predicted class by analysing several key situations. We examined high-confidence (>0.95) correct classifications for 'Healthy', 'Obvious signs', and 'Occluded' images to confirm the model's confident decisions were based on biologically relevant features. We also investigated correctly identified 'Mild signs' cases where the model exhibited lower confidence (0.5-0.75). Finally, we focused on divergent classifications: instances where the model's prediction differed from the expert label, such as when 'Healthy' animals were labelled as 'Disease' or vice versa.

It is important to recognise that expert labels are not an absolute ground truth, as no physical verification of disease status was performed. Therefore, these divergent classifications represent a disagreement between the model and the human expert, not necessarily a model error. By overlaying heatmaps onto the original images, we could qualitatively assess the features that led to a divergent prediction, providing deeper insight into the model's decision-making process.

6.3 Results

6.3.1 Supervised Classification of Rumpwear Status

The test set performance of the four supervised EfficientNetV2S model configurations is detailed in Table 6.2. All models reached 93.9%–95.2% accuracy and weighted average F1-score, confirming the architecture's suitability.

Table 6.2. Performance metrics for four EfficientNetV2S model configurations on the rumpwear test set. ImageNet_fr_ft: Pre-trained on ImageNet, frozen base then fine-tuned. Wildlife_fr_ft: Pre-trained on custom wildlife dataset, frozen base then fine-tuned. ImageNet_ft: Pre-trained on ImageNet, fine-tuned only. Wildlife_ft: Pre-trained on custom wildlife dataset, fine-tuned only. Best values for each metric are shown in bold.

Metric	**ImageNet_fr_ft**	**Wildlife_fr_ft**	**ImageNet_ft**	**Wildlife_ft**
Accuracy	0.94	0.939	0.948	**0.952**
F1-score (Healthy)	0.95	0.949	**0.958**	**0.958**

F1-score (Disease)	0.903	0.904	**0.927**	0.926
F1-score (Occluded)	0.953	0.949	0.945	**0.96**
Macro avg F1-score	0.935	0.934	0.943	**0.948**
Macro avg Precision	0.929	0.929	**0.948**	0.947
Macro avg Recall	0.943	0.939	0.939	**0.95**
Precision (Healthy)	**0.969**	0.964	0.946	0.962
Precision (Disease)	0.863	0.881	**0.939**	0.913
Precision (Occluded)	0.955	0.942	0.959	0.965
Recall (Healthy)	0.932	0.935	**0.97**	0.955
Recall (Disease)	**0.946**	0.928	0.916	0.94
Recall (Occluded)	0.951	**0.956**	0.931	**0.956**
Weighted avg F1-score	0.941	0.939	0.948	**0.952**
Weighted avg Precision	0.942	0.94	0.948	**0.952**
Weighted avg Recall	0.94	0.939	0.948	**0.952**

The Wildlife_ft model, which used custom wildlife-specific pre-trained weights and was fine-tuned directly, demonstrated the highest overall accuracy (0.952) and the best weighted-average F1-score (0.952). Comparing pre-training sources, models initialised with custom wildlife weights (Wildlife_ft) showed a slight advantage over their ImageNet counterparts (ImageNet_ft) in overall performance. Regarding the training strategy, directly fine-tuning

the models (_ft) consistently resulted in better performance than including an initial frozen base phase (_fr_ft).

For the crucial ‘Disease’ class, the ImageNet_ft configuration achieved the highest F1-score (0.927) and precision (0.939), making it the most reliable for positive predictions. The highest recall (0.946), indicating the best ability to identify all true disease cases, was achieved by the ImageNet_fr_ft model.

All four models showed very similar classification performance, so the choice of final model was driven by calibration rather than accuracy alone. Among the two strongest candidates, the ImageNet_ft model performed best on the ‘Disease’ class, but the Wildlife_ft model achieved the lowest validation loss, indicating better overall calibration. Given our focus on using model probabilities as a proxy for disease severity, we selected the Wildlife_ft configuration, accepting a small reduction in ‘Disease’ precision as a trade-off for more reliable probability estimates. All subsequent analyses therefore use Wildlife_ft as the base model.

6.3.2 Semi-Supervised Classification of Rumpwear Status

We investigated whether a semi-supervised learning (SSL) protocol could improve upon the supervised model's performance. The results were mixed: while the final SSL model increased recall for the ‘Disease’ class, this gain came at the cost of lower precision and overall accuracy. The detailed performance metrics are provided in the supplementary material (Supplementary Table 13).

6.3.3 Assessing Model Output as a Proxy for Disease-Severity

We examined model confidence across all datasets. For images in the test set, the model generally assigned high probabilities to its predicted class (Supplementary Figure 11). A similar high-confidence pattern was observed in the predictions made on the unlabelled dataset during the semi-supervised learning protocol (Supplementary Figure 12).

The model's predicted probability for the 'Disease' class (P(Disease)) increased systematically with expert-assigned disease severity (Figure 6.2). A Kruskal-Wallis test confirmed a highly significant difference in the P(Disease) distributions across the 'Healthy', 'Mild Signs', and 'Obvious Signs' groups ($H = 348.29$, $p < 0.001$). Post-hoc

analysis using Dunn's test showed that all three groups were statistically distinct from one another. The separation was most pronounced between the 'Healthy' (mean = 0.09) and 'Obvious Signs' (mean = 0.96) groups ($p < 0.001$). The 'Mild Signs' group (mean = 0.75) was also clearly distinguished from both the 'Healthy' group ($p < 0.001$) and, more subtly but still significantly, from the 'Obvious Signs' group ($p = 0.018$). This demonstrates a clear ordinal relationship where the model's confidence in its 'Disease' prediction aligns directly with the visually assessed severity of the condition.

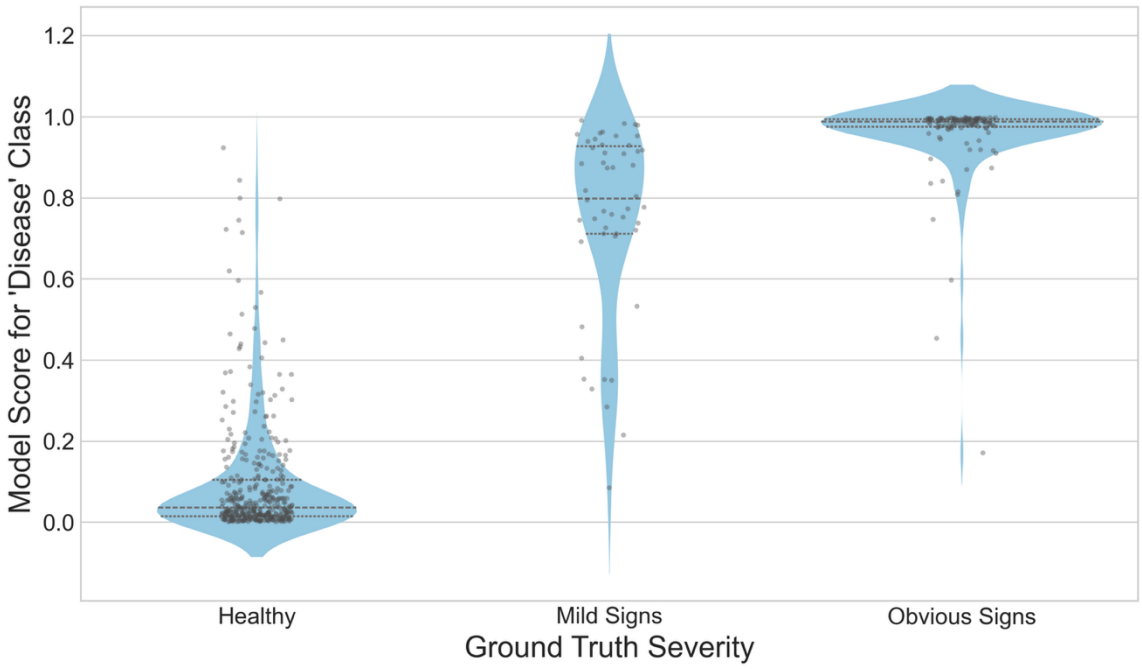


Figure 6.2. Violin plots showing the distribution of model-predicted disease probability (P(Disease)) for test set images, grouped by their expert-assigned ground truth label.

When visualised in a two-dimensional probability space, the 'Healthy', 'Mild Signs', and 'Obvious Signs' categories formed distinct regions (Supplementary Figure 13). 'Healthy' images clustered where P(Healthy) was high and P(Disease) was low, while 'Obvious Signs' clustered in the opposite corner. The 'Mild Signs' images were primarily located in the space between these two groups.

The separation of severity classes was also evident in dimensionality reduction visualisations of the model's feature embeddings, projections showed moderate clustering. The UMAP projection showed visually distinct clustering (Figure 6.3) with a silhouette score of 0.373, while the PCA and t-SNE projections resulted in similar scores of 0.372 and 0.298, respectively (Supplementary Figure 14 and Supplementary Figure 15). The first two principal components of the PCA on these bottleneck features captured 70% of the variance. These silhouette scores indicate that while some structure is present, the severity classes show considerable overlap within the bottleneck feature space. The separability was stronger in the model's final three-dimensional softmax probability outputs. Here, a PCA projection achieved a silhouette score of 0.648 with its first two components explaining nearly 100% of the variance (Supplementary Figure 16). The corresponding t-SNE and UMAP projections had lower scores of 0.342 and 0.255, respectively (Supplementary Figure 17 and Supplementary Figure 18).

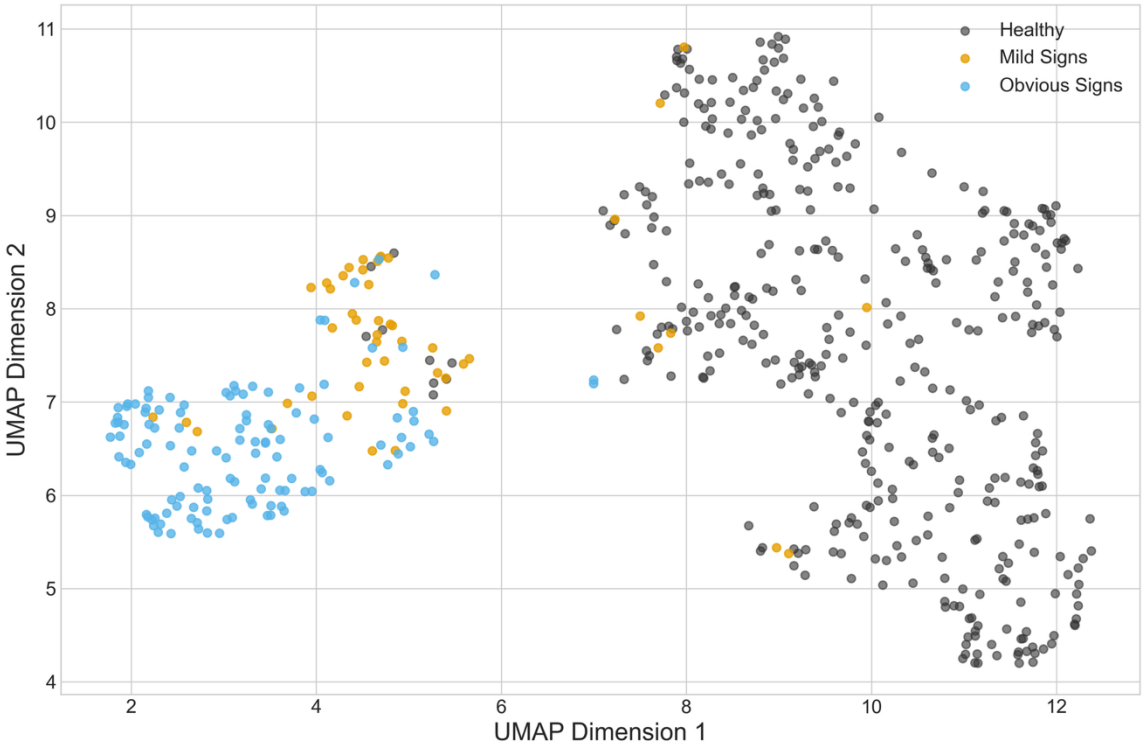


Figure 6.3 UMAP projection of the 256-dimensional feature embeddings extracted from the model's 'compression_bottleneck' layer for each image in the non-occluded test set. Points are coloured by their expert-assigned ground truth label.

The reliability of the model's confidence score as a severity proxy was confirmed with a detailed calibration analysis (Figure 6.4). The model achieved a low Brier score of 0.034 and a small ECE of 0.046.

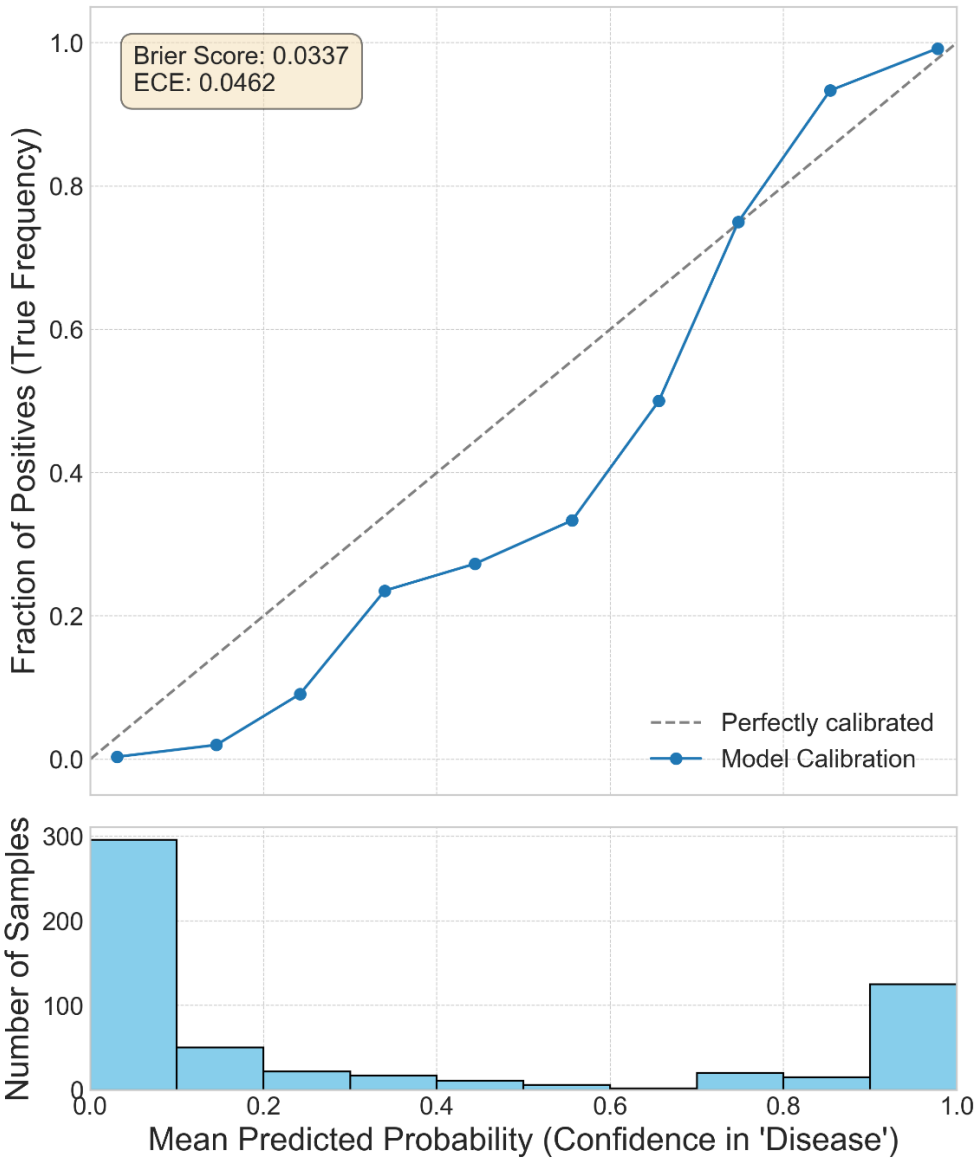


Figure 6.4. Calibration analysis of the model's 'Disease' class probability on the non-occluded test set. The top panel compares the model's predicted probability (blue line) to a perfectly calibrated model (dashed line). The model's Brier score and Expected Calibration Error (ECE) are shown in the top-left corner. The bottom panel shows a histogram of these predicted probabilities.

6.3.4 Model Interpretability: Identifying Key Image Regions

A quantitative analysis of the Grad-CAM heatmaps for all non-occluded test images showed a strong link between the model's focus and its performance. Correct classifications were highly associated with cohesive heatmaps (94.0%), whereas this figure was much lower for divergent classifications (60.7%). Heatmap cohesion also correlated directly with model confidence: 99.7% of high-confidence predictions (>0.95) had cohesive heatmaps, compared to only 67.1% for low-confidence predictions (<0.75).

The visual outputs in Figure 6.5 illustrate these quantitative patterns. For high-confidence 'Healthy' (Figure 6.5A) and 'Obvious Signs' (Figure 6.5B) classifications, the model's attention was strongly and cohesively centred on the rump. When classifying an 'Occluded' animal (Figure 6.5C), the model's focus was appropriately on the possum's head and face. Additional examples for each scenario are provided in Supplementary Figure 19 to Supplementary Figure 24.

In contrast, visualisations for uncertain or divergent classifications often showed a more fragmented activation pattern. This was evident even for some correctly identified, but challenging, 'Mild Signs' images, which had a lower proportion of cohesive heatmaps (88.0%) than 'Obvious Signs' images (98.3%) (Figure 6.5D). This fragmented pattern was most common in cases of divergent classification. When a healthy animal was classified as 'Disease' (Figure 6.5E), or when a diseased animal was classified as 'Healthy' (Figure 6.5F), the model's attention was often scattered and lacked a single, decisive point of focus.

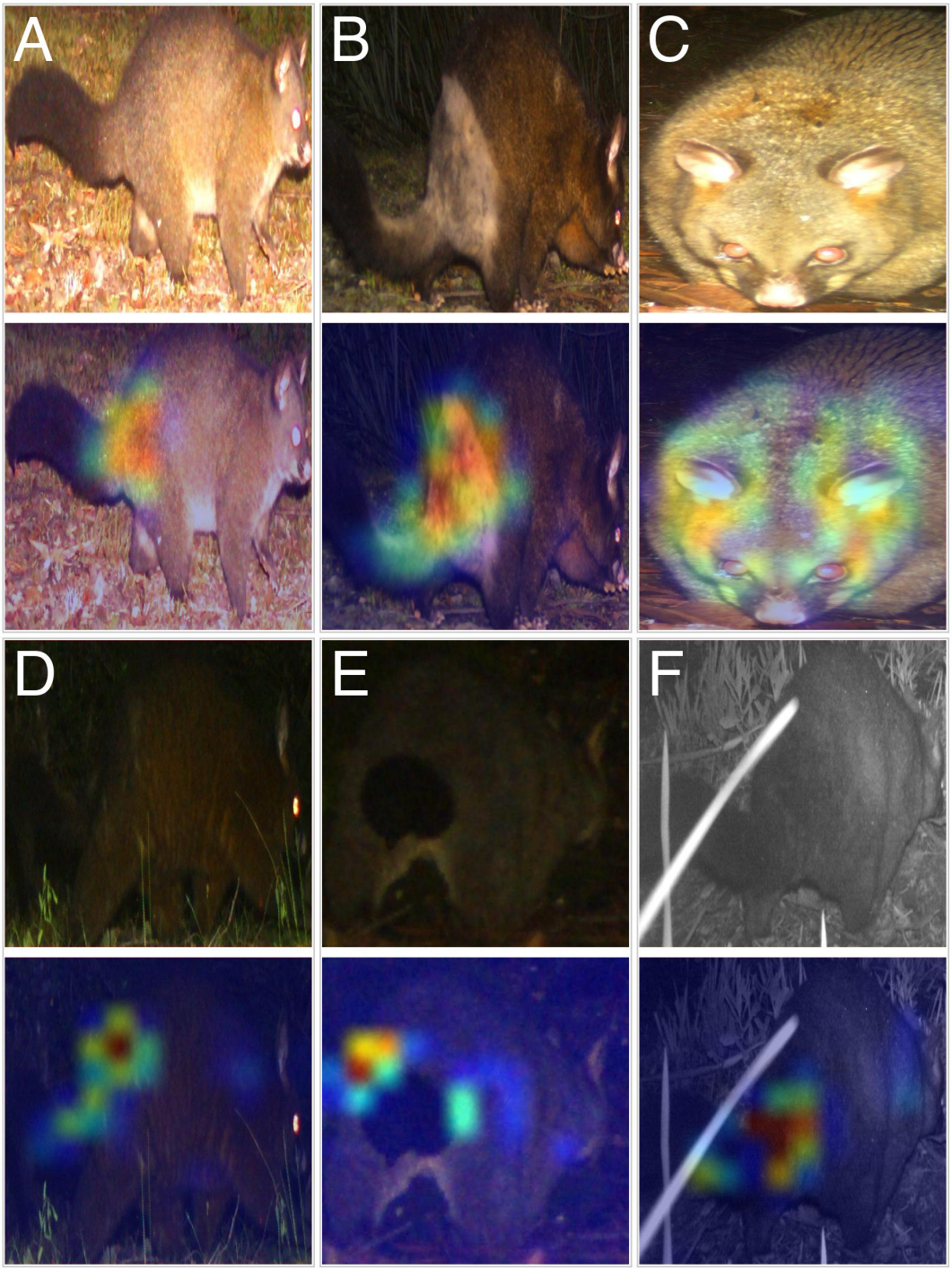


Figure 6.5. Grad-CAM visualisations of the trained wildlife_ft EfficientNetV2S model's class activation maps for representative test images. Each panel shows the original image (top) and the same image with the Grad-CAM heatmap overlay (bottom). The heatmap indicates image regions of high importance for the model's final prediction; red signifies high importance, while blue signifies low importance. The panels display six different

classification scenarios: (A) A high-confidence 'Healthy' classification. (B) A high-confidence correct ‘Disease’ classification belonging to the ‘Obvious signs’ ground truth class. (C) A high-confidence 'Occluded' classification. (D) A low-confidence, correct ‘Disease’ classification belonging to the 'Mild signs' ground truth class. (E) A divergent classification of a healthy animal as 'Disease'. (F) A divergent classification of a diseased animal as 'Healthy'.

6.4 Discussion

This study demonstrates that our pipeline automates rumpwear detection and yields a well-calibrated quantitative severity proxy. While our supervised classifier successfully learned this severity continuum, attempts to further enhance its already high performance using semi-supervised learning yielded mixed results.

Notably, the supervised model encoded the severity of rumpwear clinical signs as a continuum, mirroring its gradual visual progression (Hufschmid et al. 2010; Ringwaldt et al. 2022). This was quantitatively demonstrated by the mean P(Disease) scores on the test set, which progressed from 0.09 for 'Healthy' images, to 0.75 for 'Mild Signs', and 0.96 for 'Obvious Signs'. The model's learned feature space also reflected this: internal features showed moderate class separation (silhouette scores 0.30-0.37), while the final softmax outputs were more distinct (PCA silhouette score: 0.648). This suggests the final classification layer functions to transform the more ambiguous, overlapping features into a separable output. A detailed calibration analysis confirmed the reliability of this output as a severity proxy (Figure 6.4).

Although the model was less reliable on rare, mid-range probability predictions, it was well-calibrated overall (Brier score: 0.034). Importantly, most 'Mild Signs' cases were assigned scores in a high-confidence range (IQR: 0.71–0.93), placing them outside the most uncertain regions of the calibration curve. This validates the model's output probability not as a perfectly calibrated measure of likelihood, but as a robust correlative proxy for severity, suitable for fine-grained analysis and as a practical alternative to more complex ordinal regression models (Niu et al. 2016). Furthermore, for these intermediate cases, the model

may offer a more consistent assessment than human experts. It is plausible that the model learns to identify subtle pixel-level patterns associated with early-stage rumpwear that are difficult for the human eye to reliably classify. However, without a definitive ground truth derived from physical examination, the hypothesis that the model may exceed human performance on borderline cases cannot be verified.

Analysis of the model's decision-making process using Grad-CAM (Selvaraju et al. 2020) helped to validate its predictions. For high-confidence, correct classifications, the model focused its attention on cohesive, biologically relevant areas, specifically the rump and posterior of the animal. In contrast, for divergent classifications or low-confidence predictions, the model's focus became scattered and fragmented. This pattern suggests that the cohesion of an activation map could serve as a simple visual indicator of model uncertainty, an area that warrants further investigation. The visualisations also confirmed a key limitation: the model, much like a human expert, performed less effectively on images taken in low light or where the animal was distant. Image quality and posture strongly affect performance – a pattern seen in other camera-trap studies (Tabak et al. 2019).

Pseudo-labelling increased ‘Disease’ recall at the cost of precision, with no improvement for accuracy or F1. This outcome highlights the conditional success of SSL methods. Pseudo-labelling is typically most effective when a small expert-labelled dataset is supplemented by a much larger volume of unlabelled data, particularly when the initial supervised model's performance is not yet saturated (Sohn et al. 2020). The high accuracy of our supervised model meant the primary risk was confirmation bias, where the model iteratively amplifies its own few prediction errors (or divergences from the human labels). This process can introduce noise that negates the benefit of using more data (Arazo et al. 2020). Expanding the expert-labelled dataset would likely provide clearer gains than applying this semi-supervised approach.

Our work is a methodological advance, building upon previous studies that documented rumpwear epidemiology through manual annotation (e.g. Ringwaldt et al. 2022). It also complements the work of Ringwaldt et al. (2025), who first applied a CNN to much of the same data for broad-scale presence/absence mapping. Our study extends this foundation by demonstrating that a deep learning model can move beyond simple classification. By

validating the model's output probability as a proxy for a nuanced *severity* index, we provide a tool for more fine-grained eco-epidemiological analysis. This high-throughput, automated scoring overcomes the primary bottleneck of manual data processing, enabling analysis on a scale that would otherwise be infeasible. This tool does not replace the need for ecological expertise, but its use should give researchers more time to focus on interpreting patterns rather than labelling images. A caveat is that our model was developed using a single camera-trap network; its performance in other ecological contexts or with different camera setups would require further validation.

Looking forward, the validated supervised model can be applied to the full, unlabelled camera trap image archive from the University of Tasmania's camera network to conduct landscape-level epidemiological studies. More broadly, the methodological framework presented here can be adapted to other camera trap networks and wildlife disease systems. This study is also a starting point for further technical refinement. For instance, while our model's probabilities are well-calibrated, they could be further improved by applying post-hoc calibration methods. Temperature scaling (Guo et al. 2017), a technique for correcting for model overconfidence, could be a logical next step to fine-tune the reliability of the output probabilities before their use in downstream epidemiological models. From a data perspective, expanding the labelled dataset remains a key avenue for improving model robustness and evaluation.

6.5 Conclusion

We show that deep learning can support wildlife disease surveillance by detecting externally visible clinical signs and by providing a quantitative proxy for disease severity. Using rumpwear in common brushtail possums as a case study, our EfficientNetV2S-based pipeline achieved high classification performance on held-out test images and produced biologically plausible attention patterns in Grad-CAM visualisations.

Beyond presence/absence detection, the model's continuous output for the 'Disease' class increased systematically with expert-assigned severity categories ('Mild' to 'Obvious'), supporting its use as a practical severity index for large image collections. This provides a scalable alternative to manual grading and a transferable framework for other wildlife conditions with gradable, externally visible signs. When combined with metadata (date and

location), such severity scores can support finer-scale analyses of wildlife disease dynamics and inform targeted follow-up monitoring.

CHAPTER 7 :

GENERAL DISCUSSION

7.1 Synthesis of Key Findings

This thesis investigated the potential and challenges of using computer vision (CV) to classify visible signs of disease in wildlife from digital images. The research moved from a broad assessment of disease suitability and data availability to the empirical development and evaluation of CV models for specific case studies. This multi-faceted approach addresses key knowledge gaps in using modern technology for wildlife health surveillance.

Question 1: Which wildlife diseases, particularly those subject to existing monitoring efforts, present with externally visible signs suitable for detection via computer vision? Chapter 3 established the scope for CV-based monitoring by analysing the World Organisation for Animal Health (WOAH) disease database. The analysis showed that many animal diseases manifest with externally visible signs. Overall, 137 of the 204 unique infectious diseases (67%) and 162 of 250 disease-host associations (65%) presented with visible signs. The relevance to wildlife surveillance was clear: 85% of these visible disease-host units (138 of 162) were associated with wildlife hosts. The most common signs were lesions, swelling or oedema, and colour changes. This assessment confirmed that a significant number of existing and emerging disease concerns present observable signs, which supports using visual data to augment current surveillance methods.

Question 2: What is the availability, quality, and ecological information content of wildlife imagery, particularly for fish species (salmonids), accessible from online digital repositories, and what is the prevalence of visible disease signs within these data?

This question was addressed across two chapters. An initial broad assessment in Chapter 3 explored image availability on Flickr for the diseases in our database. This analysis showed a trend for diseases with visible signs to have more associated images than those without visible signs (a median count nearly three times higher), though this result was not statistically significant ($p=0.17$), likely due to the inherent noise in such a broad search. To investigate image availability and content in more detail, Chapter 4 then provided a critical evaluation of online digital repositories as sources of wildlife imagery, focusing on salmonids as a case study for data availability, quality, and content. Our comprehensive analysis of nearly 70,000 images revealed these platforms, especially iNaturalist, offer a vast and

rapidly expanding visual dataset with considerable spatial and temporal reach. While image quality and metadata completeness varied notably across platforms, with iNaturalist contributing a large and rapidly growing collection of high-sharpness images and GBIF ensuring metadata consistency, the data were affirmed as robust for distribution and health indicator analyses. A key finding was the consistent prevalence of 14-18% of images showing visible signs of disease or physical damage across the platforms. This highlights a significant, previously underused resource for broad-scale wildlife health surveillance. The chapter also identified practical limitations, including restrictive "All Rights Reserved" licenses (mainly on Flickr), the absence of standard size references, and temporal biases in older data.

Question 3: How effectively can deep learning algorithms be developed and trained to classify specific, visible signs of disease in wildlife images, using *Saprolegnia spp.* in salmonids as a case study, and how does dataset composition influence model performance?

Chapter 5 provided empirical evidence for the effectiveness of deep learning algorithms in classifying visible signs of *Saprolegnia* spp. infection in salmonids. All developed models consistently outperformed random classification, with EfficientNetV2S demonstrating superior performance (macro-average F1-score of 0.920 on the *Salmo* genus dataset) likely because it can process higher-resolution images. This allows it to discern subtle yet diagnostically important features, such as the fine, filamentous texture of early-stage fungal hyphae or slight skin discolouration, that would be lost in lower-resolution inputs. This work showed the potential for rapid, large-scale disease screening, offering a valuable tool for early outbreak detection and more targeted application of confirmatory diagnostics, despite not replacing molecular diagnosis. A critical contribution was the adherence to rigorous validation protocols, which provided a more realistic assessment of real-world applicability compared to prior studies.

The influence of dataset composition and image characteristics on model performance was highlighted across both the salmonid (Chapter 5) and possum (Chapter 6) case studies. Chapter 5 showed that performance was highly dependent on the taxonomic specificity of the dataset; the precision of the best model dropped from 0.858 on the *Salmo*-specific

dataset to just 0.462 on the *Oncorhynchus*-specific dataset. The better performance on the *Salmo* dataset was likely due to it being more balanced, whereas data imbalance in other datasets contributed to the overprediction of the minority (disease) class. Beyond simple data imbalance, this discrepancy may also reflect other confounding factors, such as subtle inter-genus differences in the visual presentation of saprolegniosis or potential systematic variations in image quality between the datasets, making classification inherently more challenging for the *Oncorhynchus* genus. The influence of image quality was also apparent: performance was affected by water surface reflections in the aquatic environment (Chapter 5), and by factors like low light or subject distance in the terrestrial camera-trap images (Chapter 6).

Question 4: Can the outputs of computer vision classification models serve as a proxy for disease severity assessment, and can semi-supervised learning approaches help mitigate labelled data limitations in this context?

Chapter 6 expanded the application of CV beyond binary classification to assess disease severity, using brushtail possum rumpwear as a case study. The underlying supervised model achieved high performance , with results showing that pre-training on a custom wildlife dataset from the same camera network provided a slight advantage over using standard ImageNet weights, reaching an accuracy of up to 95.2% and a 'Disease' class F1-score of 0.927. Trained only on broad 'Healthy', 'Disease' and ‘Occluded’ labels, it produced an output probability that served as a robust proxy for disease severity. This was demonstrated by the mean model outputs for the ‘Disease’ class, which progressed from 0.09 for 'Healthy' images, to 0.75 for 'Mild Signs', and 0.96 for 'Obvious Signs'. The model was well-calibrated for this purpose (Brier score: 0.034), providing a high-throughput, automated scoring method that overcomes manual processing bottlenecks. Importantly, this method moves beyond coarse, categorical health scores by generating a continuous, quantitative variable. This innovation could open the door to more powerful statistical analyses of disease dynamics, such as modelling subtle population-level shifts in mean disease severity over time or in response to management interventions. The chapter also explored semi-supervised learning to address data limitations; this yielded mixed results, increasing recall at the cost of precision, which suggests a possible ceiling effect for the already high-performing model. Finally, model interpretation using Grad-CAM validated that

the model focused on relevant anatomical features, confirming its utility while also highlighting its limitations with poor quality images. The interpretability was quantitatively validated: 94% of correct classifications produced cohesive heatmaps, compared to only 61% of incorrect classifications. This method moves beyond the subjective visual assessment of heatmaps towards a quantitative, data-driven measure of when the model's decisions can be trusted.

7.2 Overall Implication and Contributions to Wildlife Disease Surveillance

The findings of this thesis advance the practical application of computer vision (CV) for wildlife disease surveillance, offering a complement to traditional monitoring methods. By investigating the entire pipeline from disease suitability to model development, this research shows the practical utility of CV for ecological science and wildlife health management.

This thesis shows that CV offers a way to address existing gaps in wildlife disease surveillance. Traditional methods often face limitations in scale, timeliness, and geographical coverage (Delgado et al. 2023; Barroso et al. 2024). Notably, the methods developed in this thesis align directly with the most recent international guidance, which formally recognises "non-biological samples", including visual observations from camera traps, as a key component of modern wildlife surveillance (WOAH and IUCN 2024). Our work addresses these challenges by using automated image analysis for non-invasive, scalable monitoring. The analysis of online digital repositories (Chapter 4), particularly the high volume of observations on platforms like iNaturalist, highlights a large data stream that can be used for broad-scale surveillance without needing to physically capture animals. This automated, non-invasive approach provides a real alternative to the costly, logistically intensive, and often stressful capture-recapture programs that are frequently the only other means of assessing the health of free-ranging populations, thereby reducing costs, complexity, and impacts on animal welfare.

The case studies in Chapters 5 and 6 show the effectiveness of deep learning models in extracting meaningful ecological information from these image sources. The successful

classification of *Saprolegnia* spp. in salmonids (Chapter 5) provides a method for rapid, large-scale screening, which can act as an early warning system. This approach allows for more targeted application of costly confirmatory diagnostics by flagging potential outbreaks. Similarly, the ability of CV to serve as a proxy for disease severity, as shown with possum rumpwear (Chapter 6), extends beyond simple presence or absence detection. This offers a high-throughput, automated scoring method that overcomes the bottleneck of manual processing, enabling fine-grained ecological analysis of disease progression across populations.

By integrating these CV-derived health classifications with spatio-temporal metadata, this research opens new avenues for epidemiological analyses. These include mapping disease distributions, tracking spread, and investigating associations with environmental drivers (Toivonen et al. 2019), improving our understanding of disease dynamics in wild populations. The insights from this thesis show the potential for CV to extend the reach and efficiency of wildlife disease surveillance, enabling more timely and comprehensive responses to emerging health threats.

7.3 Broader Challenges and Limitations

Despite the potential shown in this thesis, the successful and widespread application of CV for wildlife disease surveillance faces several substantial and interconnected challenges. These limitations, identified across various stages of the research, require ongoing attention and methodological innovation.

A fundamental and pervasive challenge lies in data availability and quality. While online repositories offer vast quantities of imagery, images explicitly showing clear, specific disease signs remain inherently scarce (Green et al. 2020). This rarity, coupled with potential observer bias against photographing unhealthy animals, frequently results in highly imbalanced datasets, posing a significant hurdle for training robust and unbiased machine learning models (Chapter 2; Chapter 5). However, the opposite effect, a ‘novelty bias’ must also be considered, where citizen scientists are disproportionately motivated to photograph and upload an unusually sick-looking animal precisely because of its rarity (Edwards et al. 2021). This competing bias could, in certain contexts, lead to an artificial inflation of prevalence estimates derived from these platforms. Acknowledging this duality of potential

under- and over-reporting is critical for interpreting health data from opportunistic sources. Furthermore, the opportunistic nature of much online imagery leads to immense variability in quality, including inconsistent lighting, animal pose, distance, occlusion, and complex backgrounds (Cunha et al. 2023; Chapter 2; Chapter 4; Chapter 5). This is particularly acute in aquatic environments, where factors like water turbidity, surface reflections, and rapid subject movement can obscure subtle disease features (Chapter 5).

Data labelling and ground truth validation present a further challenge. Reliably identifying visible disease signs can be ambiguous, as visual cues may mimic natural variation or injuries. This ambiguity complicates the creation of high-quality, consistent ground truth labels needed for supervised learning (Murray et al. 2021; Chapter 2; Chapter 5). The ideal of validating model outputs against 'gold standard' clinical diagnoses is rarely feasible in free-ranging wildlife populations, which highlights the importance of rigorous protocol design and expert involvement in the labelling process (Chapter 2). A key limitation of relying on visual signs is the inability to investigate the underlying cause of diseases. This challenge of differential diagnosis is particularly acute, as some conditions have similar visual presentations despite different causes. Consequently, image-based surveillance may underestimate true disease prevalence or misattribute signs, reinforcing its role as a screening tool rather than a definitive diagnostic (Ringwaldt et al. 2022).

Methodological complexities in model development also present limitations. There is a risk of models learning spurious correlations, associating disease labels with irrelevant background elements rather than the actual pathological features, which severely limits their generalisation capabilities to new environments or conditions (Miao et al. 2019; Chapter 2). While transfer learning aids in mitigating data requirements, and advanced techniques like semi-supervised learning hold promise, our findings (Chapter 6) indicate that for already high-performing models or datasets, these methods may offer only marginal gains. For such models, semi-supervised learning runs the risk of introducing noise from the unlabelled set that can degrade precision, highlighting that more complex methods do not guarantee improved performance. Emerging Vision-Language Models, while promising, currently struggle with the fine-grained distinctions and specialised terminology required for accurate disease identification (Vendrow et al. 2024)

Beyond the technical aspects, practical and operational constraints must also be considered. Developing deep learning workflows typically demands significant computational power and specialised technical expertise, which can be barriers for many ecological research groups (Ditria et al. 2020; Vidal et al. 2021; Chapter 2; Chapter 5). Furthermore, the licensing restrictions prevalent on certain online platforms (e.g., "All Rights Reserved" on Flickr) significantly curtail widespread data reuse, impeding the development and dissemination of AI models and curated datasets (Chapter 4). The increasing capabilities of AI-driven image editing features in consumer devices also pose a nascent but significant threat to the scientific integrity of online visual data, as algorithmic alterations may create convincing but false representations that are not transparently documented (Chapter 4). Lastly, evolving fish welfare guidelines advocating for in-water photography, while ethically laudable, may inadvertently reduce image clarity and visibility of key features, presenting a new challenge for data quality and the scientific utility of angler-sourced images (Chapter 4).

Collectively, these challenges underscore that while CV offers immense potential, its effective application in wildlife disease surveillance requires continuous innovation in data acquisition, meticulous curation, robust methodological development, and careful consideration of the inherent biases and limitations of opportunistic visual data and the very nature of visual disease signs themselves.

7.4 Future Work

Building upon the foundations and insights generated by this thesis, several key directions emerge for future research to further advance the application of computer vision (CV) in wildlife disease surveillance. These directions encompass improvements in data acquisition, methodological refinements, enhanced validation, and broader deployment strategies, collectively pushing the boundaries of landscape epidemiology.

7.4.1 Data Acquisition and Curation

A primary focus for future work should be on strategies to overcome the persistent challenge of data scarcity and class imbalance. This includes exploring and promoting standardised imaging protocols and metadata structures, particularly for citizen science contributions (Chapter 4), to ensure higher quality and more consistent data. Engaging directly with contributor communities, such as anglers, to educate them on best practices

for capturing scientifically valuable images of both healthy and unhealthy-looking animals could significantly improve dataset representativeness and reduce bias (Chapter 5). Maximizing the utility of all available image data, including those with incomplete metadata or unverified finer taxonomic labels, for training robust AI models, remains a critical future direction (Chapter 4). Further research into multi-source data integration, beyond the platforms explored here, could provide a more comprehensive picture.

7.4.2 Methodological Advancements and New Frontiers

Future research should continue to explore advanced learning paradigms to mitigate labelled data limitations. This includes deeper investigation into semi-supervised and self-supervised learning techniques to leverage potentially abundant unlabelled imagery (Chapter 2, Section 2.6.2; Chapter 6). Further research into few-shot learning could also be highly beneficial for identifying rare or emerging disease signs, allowing models to generalise from very limited examples (Chapter 2, Section 2.3). Developing more robust models for challenging environmental conditions, particularly for underwater imagery in aquatic systems, is crucial; this could involve incorporating explicit fish detection tools, such as 'megafishdetector' (Yang et al. 2023), and advanced image pre-processing techniques to reduce reflections and enhance clarity (Chapter 5). Furthermore, a powerful methodological frontier involves integrating the model disease classifications with external environmental datasets to test specific ecological hypotheses. This would enable researchers to investigate long-standing ecological questions, such as whether outbreaks are positively correlated with periods of low flow and high water temperature, transforming the CV tool from a simple detector into an engine for epidemiological discovery. Exploring hierarchical models that incorporate taxonomic data could also improve classification accuracy by leveraging broader biological relationships (Elhamod et al. 2022).

A significant "new frontier" lies in the further development and application of Vision-Language Models (VLMs) (Radford et al. 2021; Vendrow et al. 2024). These models hold immense potential for more flexible data querying using natural language and for open-set recognition; the ability to identify samples that do not belong to known categories, which is highly relevant for detecting previously unseen disease conditions or anomalies (Chapter 2). For instance, this could enable a system trained only on known diseases to flag an image of a possum with a novel condition, such as unusual skin nodules, as an 'unknown anomaly',

triggering an alert for expert veterinary review. While currently facing limitations with fine-grained distinctions, their continued refinement could revolutionize surveillance by allowing for the detection of novel disease signs not explicitly trained for. Finally, a challenging but valuable research avenue is the development of AI that can infer morphometric data from images lacking an explicit size reference. This might be achievable by leveraging the fixed geometry of camera traps to estimate distance, training models to use common background objects as opportunistic scales, or by learning consistent body-part ratios to assess relative animal condition (Chapter 4).

7.4.3 Model Evaluation and Validation

To ensure the practical utility of CV models, future work must focus on rigorous evaluation beyond standard computer vision metrics. This includes developing methods to assess whether model performance translates directly into accurate ecological variables like prevalence estimates and understanding the sensitivity of these estimates to various data biases (Pantazis et al. 2024). Refining model interpretation techniques, such as saliency maps, to serve as visual indicators of model uncertainty, or to verify that models are focusing on relevant pathological features rather than spurious background cues, is also important (Chapter 6). Addressing the emerging concern of AI-driven image editing in consumer devices will be critical, necessitating new standards for transparent documentation of alterations to maintain scientific integrity of visual data from public sources (Coalition for Content Provenance and Authenticity 2025, Chapter 4).

7.4.4 Deployment and Broader Impact

Efforts should continue towards developing highly efficient models suitable for deployment on edge devices (Tuia et al. 2022). In an ecological context, these are field-based hardware, such as smart camera traps or drones, that process data locally rather than requiring it to be sent to a remote server. This local processing could enable real-time disease detection in the field. This would facilitate direct, actionable insights for on-the-ground management. Fostering interdisciplinary collaboration between ecologists, veterinarians, and computer scientists remains paramount to bridge expertise gaps and ensure successful implementation of these advanced tools (Weinstein 2018). Finally, advocating for and promoting open licensing policies across all online platforms would substantially enhance

the accessibility and reproducibility of research (Wilkinson et al. 2016), bolstering global efforts to address pressing conservation and health challenges (Chapter 4). This includes encouraging data infrastructure that supports standardised protocols and ethical data sharing of sensitive wildlife health information (Tulloch et al. 2018; WOAH and IUCN 2024).

7.5 Conclusion

This thesis has shown the potential of computer vision for advancing wildlife disease surveillance. Through a multi-faceted research design, it identified a broad range of visually apparent wildlife diseases, critically assessed online image repositories as novel data sources, and confirmed the effectiveness of deep learning models for both disease classification and severity assessment.

By using opportunistic visual data, this work has contributed to bridging knowledge gaps in wildlife health monitoring. It shows how CV can overcome traditional logistical constraints, enabling non-invasive, scalable, and efficient surveillance. While acknowledging challenges such as data scarcity, the variability of field-captured imagery, and the limitations of visual-only diagnosis, this research has also identified practical strategies and promising avenues for future work.

The work presented here provides a framework and evidence for integrating computer vision into applied wildlife health monitoring. It moves beyond simply offering a new tool; it demonstrates how to transform the casual observation into a quantitative data point, fundamentally enhancing our ability to understand disease dynamics and support timely conservation responses on a global scale.

REFERENCES

Ahmed, M.S. 2024. SalmonScan: A Novel Image Dataset for Machine Learning and Deep Learning Analysis in Fish Disease Detection in Aquaculture. 3. Available at: https://data.mendeley.com/datasets/x3fz2nfm4w/3 [Accessed: 8 January 2025].

Ahmed, M.S., Aurpa, T.T. and Azad, Md.A.K. 2022. Fish Disease Detection Using Image Based Machine Learning Technique in Aquaculture. *Journal of King Saud University - Computer and Information Sciences* 34(8, Part A), pp. 5170–5182. doi: 10.1016/j.jksuci.2021.05.003.

Alkhawaldeh, I.M., Albalkhi, I. and Naswhan, A.J. 2023. Challenges and limitations of synthetic minority oversampling techniques in machine learning. *World Journal of Methodology* 13(5), pp. 373–378. doi: 10.5662/wjm.v13.i5.373%5D.

Alom, M.Z. et al. 2019. A State-of-the-Art Survey on Deep Learning Theory and Architectures. *Electronics* 8(3). doi: 10.3390/electronics8030292.

Amano, T., Lamming, J.D.L. and Sutherland, W.J. 2016. Spatial Gaps in Global Biodiversity Information and the Role of Citizen Science. *BioScience* 66(5), pp. 393–400. doi: 10.1093/biosci/biw022.

Amara, J., Bouaziz, B. and Algergawy, A. 2017. A deep learning-based approach for banana leaf diseases classification. *Datenbanksysteme für Business, Technologie und Web (BTW 2017)-Workshopband*.

Antonovics, J. et al. 2017. The evolution of transmission mode. *Philosophical Transactions of the Royal Society B: Biological Sciences* 372(1719), p. 20160083. doi: 10.1098/rstb.2016.0083.

Araneda, C., Neira, R., Lam, N. and Iturra, P. 2008. Salmonids. *Genome mapping and genomics in fishes and aquatic animals*, pp. 1–43.

Aravamuthan, S., Cernek, P., Anklam, K. and Döpfer, D. 2024. Comparative analysis of computer vision algorithms for the real-time detection of digital dermatitis in dairy cows. *Preventive Veterinary Medicine* 229, p. 106235. doi: 10.1016/j.prevetmed.2024.106235.

Arazo, E., Ortego, D., Albert, P., O'Connor, N.E. and McGuinness, K. 2020. Pseudo-Labeling and Confirmation Bias in Deep Semi-Supervised Learning. In: *2020 International Joint Conference on Neural Networks (IJCNN)*. pp. 1–8. Available at: https://ieeexplore.ieee.org/document/9207304 [Accessed: 7 July 2025].

Arlinghaus, R. et al. 2020. Global Participation in and Public Attitudes Toward Recreational Fishing: International Perspectives and Developments. *Reviews in Fisheries Science & Aquaculture* 29(1), pp. 58–95. doi: 10.1080/23308249.2020.1782340.

Arshad, B., Barthelemy, J., Pilton, E. and Perez, P. 2020. Where is my Deer?-Wildlife Tracking And Counting via Edge Computing And Deep Learning. In: *2020 IEEE SENSORS*. pp. 1–4. doi: 10.1109/SENSORS47125.2020.9278802.

Artstein, R. 2017. Inter-annotator Agreement. In: Ide, N. and Pustejovsky, J. eds *Handbook of Linguistic Annotation*. Dordrecht: Springer Netherlands, pp. 297–313.

Atlas, W.I. et al. 2023. Wild salmon enumeration and monitoring using deep learning empowered detection and tracking. *Frontiers in Marine Science* 10. Available at: https://www.frontiersin.org/journals/marine-science/articles/10.3389/fmars.2023.1200408/full [Accessed: 19 February 2025].

August, T., Fox, R., Roy, D.B. and Pocock, M.J.O. 2020a. Data-derived metrics describing the behaviour of field-based citizen scientists provide insights for project design and modelling bias. *Scientific Reports* 10(1), p. 11009. doi: 10.1038/s41598-020-67658-3.

August, T.A., Pescott, O.L., Joly, A. and Bonnet, P. 2020b. AI Naturalists Might Hold the Key to Unlocking Biodiversity Data in Social Media Imagery. *Patterns* 1(7), p. 100116. doi: https://doi.org/10.1016/j.patter.2020.100116.

Barbedo, J.G.A., Gomes, C.C.G., Cardoso, F.F., Domingues, R., Ramos, J.V. and McManus, C.M. 2017. The use of infrared images to detect ticks in cattle and proposal of an algorithm for quantifying the infestation. *Veterinary Parasitology* 235, pp. 106–112. doi: https://doi.org/10.1016/j.vetpar.2017.01.020.

Barcina-Blanco, M., Lobo, J.L., Garcia-Bringas, P. and Del Ser, J. 2024. Managing the unknown in machine learning: Definitions, related areas, recent advances, and prospects. *Neurocomputing* 599, p. 128073. doi: 10.1016/j.neucom.2024.128073.

Barney, S., Dlay, S., Crowe, A., Kyriazakis, I. and Leach, M. 2023. Deep learning pose estimation for multi-cattle lameness detection. *Scientific Reports* 13(1), p. 4499. doi: 10.1038/s41598-023-31297-1.

Barroso, P., López-Olvera, J.R., Kiluba wa Kiluba, T. and Gortázar, C. 2024. Overcoming the limitations of wildlife disease monitoring. *Research Directions: One Health* 2, p. e3. doi: 10.1017/one.2023.16.

Beery, S., Cole, E., Parker, J., Perona, P. and Winner, K. 2021. Species Distribution Modeling for Machine Learning Practitioners: A Review. In: *ACM SIGCAS Conference on Computing and Sustainable Societies*. Association for Computing Machinery, pp. 329–348.

Beery, S., Morris, D. and Yang, S. 2019. Efficient Pipeline for Camera Trap Image Review. Available at: https://arxiv.org/abs/1907.06772.

Beery, S., Van Horn, G. and Perona, P. 2018. Recognition in terra incognita. pp. 456–473.

Bengio, Y., Louradour, J., Collobert, R. and Weston, J. 2009. Curriculum learning. In: *Proceedings of the 26th Annual International Conference on Machine Learning*. ICML '09.

New York, NY, USA: Association for Computing Machinery, pp. 41–48. Available at: https://dl.acm.org/doi/10.1145/1553374.1553380 [Accessed: 7 July 2025].

Benjamens, S., Dhunnoo, P. and Meskó, B. 2020. The state of artificial intelligence-based FDA-approved medical devices and algorithms: an online database. *npj Digital Medicine* 3, pp. 118–118. doi: 10.1038/s41746-020-00324-0.

Berckmans, D. 2017. General introduction to precision livestock farming. *Animal Frontiers* 7(1), pp. 6–11. doi: 10.2527/af.2017.0102.

Berrada, D. 2023. *4 new Google Photos features on Pixel 8 and Pixel 8 Pro*. Available at: https://blog.google/products/photos/google-photos-features-pixel-8-pro/ [Accessed: 9 July 2025].

Bhagwat, R. and Dandawate, Y. 2021. Comprehensive Multilayer Convolutional Neural Network for Plant Disease Detection. *International Journal of Advanced Computer Science and Applications (IJACSA)* 12(1). Available at: https://thesai.org/Publications/ViewPaper?Volume=12&Issue=1&Code=IJACSA&SerialNo=25 [Accessed: 11 April 2025].

Biggs, R., Carpenter, S.R. and Brock, W.A. 2009. Turning back from the brink: Detecting an impending regime shift in time to avert it. *Proceedings of the National Academy of Sciences* 106(3), p. 826. doi: 10.1073/pnas.0811729106.

Bischoff, V., Farias, K., Menzen, J.P. and Pessin, G. 2021. Technological support for detection and prediction of plant diseases: A systematic mapping study. *Computers and Electronics in Agriculture* 181, p. 105922. doi: 10.1016/j.compag.2020.105922.

Biswas, S., Muduli, D., Islam, M.A., Kanade, A.S., Zamani, A.T., Kanade, S.P. and Parveen, N. 2024. Empirical Evaluation of Deep Learning Techniques for Fish Disease Detection in Aquaculture Systems: A Transfer Learning and Fusion-Based Approach. *IEEE access : practical innovations, open solutions* 12, pp. 176136–176154. doi: 10.1109/ACCESS.2024.3504283.

Blair, J.D., Gaynor, K.M., Palmer, M.S. and Marshall, K.E. 2024. A gentle introduction to computer vision-based specimen classification in ecological datasets. *Journal of Animal Ecology* 93(2), pp. 147–158. doi: 10.1111/1365-2656.14042.

Bochkovskiy, A., Wang, C.-Y. and Liao, H.-Y.M. 2020. YOLOv4: Optimal Speed and Accuracy of Object Detection. p. arXiv:2004.10934.

Bommasani, R. et al. 2021. On the Opportunities and Risks of Foundation Models. *arXiv e-prints*, p. arXiv:2108.07258. doi: 10.48550/arXiv.2108.07258.

Bora, M., M, M., Mathew, D.D., Das, H., Bora, D.P. and Barman, N.N. 2022. Point of care diagnostics and non-invasive sampling strategy: a review on major advances in veterinary diagnostics. *Acta Veterinaria Brno* 91(1), pp. 17–34. doi: 10.2754/avb202291010017.

Boulent, J., Foucher, S., Théau, J. and St-Charles, P.-L. 2019. Convolutional Neural Networks for the Automatic Identification of Plant Diseases. *Frontiers in Plant Science* Volume 10-2019. Available at: https://www.frontiersin.org/journals/plant-science/articles/10.3389/fpls.2019.00941.

Braga-Neto, U. 2024. *Fundamentals of Pattern Recognition and Machine Learning*. Cham: Springer International Publishing. Available at: https://link.springer.com/10.1007/978-3-031-60950-3 [Accessed: 17 April 2025].

Brook, B.W., Buettel, J.C., van Lunteren, P., Rajmohan, P.P. and Aandahl, R.Z. 2025. MEWC: A user-friendly AI workflow for customised wildlife-image classification. *Peer Community Journal* 5. Available at: https://peercommunityjournal.org/articles/10.24072/pcjournal.565/ [Accessed: 9 June 2025].

Brown, E.D. and Williams, B.K. 2019. The potential for citizen science to produce reliable and useful information in ecology. *Conservation Biology* 33(3), pp. 561–569. doi: 10.1111/cobi.13223.

Bucko, M. and Gieger, S. 2019. *Elephant endotheliotropic herpesviruses (EEHV) (Infection with)*. World Organisation for Animal Health. Available at: https://www.woah.org/app/uploads/2022/02/elephant-endotheliotropic-herpesviruses-eehv-infection-with.pdf.

Burton, A.C. et al. 2015. REVIEW: Wildlife camera trapping: a review and recommendations for linking surveys to ecological processes. *Journal of Applied Ecology* 52(3), pp. 675–685. doi: 10.1111/1365-2664.12432.

Callaghan, C.T., Poore, A.G.B., Major, R.E., Rowley, J.J.L. and Cornwell, W.K. 2019. Optimizing future biodiversity sampling by citizen scientists. *Proceedings of the Royal Society B: Biological Sciences* 286(1912), p. 20191487. doi: 10.1098/rspb.2019.1487.

Casaer, J., Milotic, T., Liefting, Y., Desmet, P. and Jansen, P. 2019. Agouti: A platform for processing and archiving of camera trap images. In: *Biodiversity Information Science and Standards*. Pensoft Publishers, p. e46690. Available at: https://biss.pensoft.net/article/46690/ [Accessed: 20 July 2025].

Cheng, T.L. et al. 2021. The scope and severity of white-nose syndrome on hibernating bats in North America. *Conservation Biology* 35(5), pp. 1586–1597. doi: 10.1111/cobi.13739.

Chicco, D. and Jurman, G. 2020. The advantages of the Matthews correlation coefficient (MCC) over F1 score and accuracy in binary classification evaluation. *BMC Genomics* 21(1), p. 6. doi: 10.1186/s12864-019-6413-7.

Chollet, F. 2021. *Deep learning with Python*. 2nd edn. Manning Publications Co.

Christin, S., Hervet, É. and Lecomte, N. 2019. Applications for deep learning in ecology. *Methods in Ecology and Evolution* 10(10), pp. 1632–1644. doi: https://doi.org/10.1111/2041-210X.13256.

Coalition for Content Provenance and Authenticity. 2025. *Content Credentials : C2PA Technical Specification : C2PA Specifications*. Available at: https://spec.c2pa.org/specifications/specifications/2.2/specs/C2PA_Specification.html [Accessed: 9 July 2025].

Collen, B. et al. 2014. Global patterns of freshwater species diversity, threat and endemism. *Global Ecology and Biogeography* 23(1), pp. 40–51. doi: https://doi.org/10.1111/geb.12096.

Constable, P.D., Hinchcliff, K.W., Done, S.H. and Grünberg, W. eds. 2017. 1 - Clinical Examination and Making a Diagnosis. In: *Veterinary Medicine (Eleventh Edition)*. W.B. Saunders, pp. 1–28. Available at: https://www.sciencedirect.com/science/article/pii/B9780702052460000012.

Costa, S. and Lopes, I. 2022. Saprolegniosis in amphibians: An integrated overview of a fluffy killer disease. *Journal of Fungi* 8(5), p. 537. doi: 10.3390/jof8050537.

Cubuk, E.D., Zoph, B., Shlens, J. and Le, Q.V. 2020. Randaugment: Practical automated data augmentation with a reduced search space. In: *Proceedings of the IEEE/CVF conference on computer vision and pattern recognition workshops*. pp. 702–703. doi: 10.1109/CVPRW50498.2020.00359.

Cunha, F., Santos, E.M. dos and Colonna, J.G. 2023. Bag of tricks for long-tail visual recognition of animal species in camera-trap images. *Ecological Informatics* 76, p. 102060. doi: https://doi.org/10.1016/j.ecoinf.2023.102060.

Cunningham, A.A., Daszak, P. and Wood, J. 2017. One health, emerging infectious diseases and wildlife: two decades of progress? *Philosophical Transactions of the Royal Society B: Biological Sciences* 372(1725), p. 20160167. doi: 10.1098/rstb.2016.0167.

Daszak, P., Cunningham, A.A. and Hyatt, A.D. 2000. Emerging infectious diseases of wildlife--threats to biodiversity and human health. *Science (New York, N.Y.)* 287(5452), pp. 443–449. doi: 10.1126/science.287.5452.443.

Daume, S. 2016. Mining Twitter to monitor invasive alien species — An analytical framework and sample information topologies. *Ecological Informatics* 31, pp. 70–82. doi: https://doi.org/10.1016/j.ecoinf.2015.11.014.

Delgado, M. et al. 2023. Wildlife health surveillance: gaps, needs and opportunities. 42, p. 161. doi: 10.20506/rst.42.3359.

Deng, J., Dong, W., Socher, R., Li, L.-J., Li, K. and Fei-Fei, L. 2009. ImageNet: A large-scale hierarchical image database. In: *2009 IEEE Conference on Computer Vision and Pattern Recognition*. pp. 248–255. Available at: https://ieeexplore.ieee.org/document/5206848 [Accessed: 7 May 2025].

Depauw, L. et al. 2022. The use of photos to investigate ecological change. *Journal of Ecology* 110(6), pp. 1220–1236. doi: 10.1111/1365-2745.13876.

Derevnina, L. et al. 2016. Emerging oomycete threats to plants and animals. *Philosophical Transactions of the Royal Society B: Biological Sciences* 371(1709), p. 20150459. doi: 10.1098/rstb.2015.0459.

Dhondt, A.A., Tessaglia, D.L. and Slothower, R.L. 1998. Epidemic mycoplasmal conjunctivitis in house finches from eastern North America. *Journal of Wildlife Diseases* 34(2), pp. 265–280. doi: 10.7589/0090-3558-34.2.265.

Di Cecco, G.J., Barve, V., Belitz, M.W., Stucky, B.J., Guralnick, R.P. and Hurlbert, A.H. 2021. Observing the Observers: How Participants Contribute Data to iNaturalist and Implications for Biodiversity Science. *BioScience* 71(11), pp. 1179–1188. doi: 10.1093/biosci/biab093.

Dias, M.S., Tedesco, P.A., Hugueny, B., Jézéquel, C., Beauchard, O., Brosse, S. and Oberdorff, T. 2017. Anthropogenic stressors and riverine fish extinctions. *Ecological Indicators* 79, pp. 37–46. doi: https://doi.org/10.1016/j.ecolind.2017.03.053.

Ding, K., Ma, K., Wang, S. and Simoncelli, E.P. 2021. Comparison of Full-Reference Image Quality Models for Optimization of Image Processing Systems. *International Journal of Computer Vision* 129(4), pp. 1258–1281. doi: 10.1007/s11263-020-01419-7.

Ditria, E.M., Lopez-Marcano, S., Sievers, M., Jinks, E.L., Brown, C.J. and Connolly, R.M. 2020. Automating the Analysis of Fish Abundance Using Object Detection: Optimizing Animal Ecology With Deep Learning. *Frontiers in Marine Science* 7(429). doi: 10.3389/fmars.2020.00429.

Dobson, A. et al. 2006. Habitat Loss, Trophic Collapse, and the Decline of Ecosystem Services. *Ecology* 87(8), pp. 1915–1924. doi: 10.1890/0012-9658(2006)87%5B1915:HLTCAT%5D2.0.CO;2.

Dong, J., Fuentes, A., Zhou, H., Jeong, Y., Yoon, S. and Park, D.S. 2024. The impact of fine-tuning paradigms on unknown plant diseases recognition. *Scientific Reports* 14(1), p. 17900. doi: 10.1038/s41598-024-66958-2.

Dosovitskiy, A. et al. 2021. An Image is Worth 16x16 Words: Transformers for Image Recognition at Scale. Available at: http://arxiv.org/abs/2010.11929 [Accessed: 17 January 2024].

Dunbar, M.R., Johnson, S.R., Rhyan, J.C. and McCollum, M. 2009. Use of infrared thermography to detect thermographic changes in mule deer (Odocoileus hemionus) experimentally infected with foot-and-mouth disease. *Journal of Zoo and Wildlife Medicine: Official Publication of the American Association of Zoo Veterinarians* 40(2), pp. 296–301. doi: 10.1638/2008-0087.1.

Durso, A.M. et al. 2021. Citizen science and online data: Opportunities and challenges for snake ecology and action against snakebite. *Toxicon: X* 9–10, p. 100071. doi: https://doi.org/10.1016/j.toxcx.2021.100071.

Edwards, T., Jones, C.B., Perkins, S.E. and Corcoran, P. 2021. Passive citizen science: The role of social media in wildlife observations. *PLoS One* 16(8), p. e0255416. doi: https://doi.org/10.1371/journal.pone.0255416.

Elhamod, M. et al. 2022. Hierarchy-guided neural network for species classification. *Methods in Ecology and Evolution* 13(3), pp. 642–652. doi: https://doi.org/10.1111/2041-210X.13768.

Elmer, F., Kohl, Z.F., Johnson, P.T.J. and Peachey, R.B.J. 2019. Black spot syndrome in reef fishes: using archival imagery and field surveys to characterize spatial and temporal distribution in the Caribbean. *Coral Reefs* 38(6), pp. 1303–1315. doi: 10.1007/s00338-019-01843-3.

van Engelen, J.E. and Hoos, H.H. 2020. A survey on semi-supervised learning. *Machine Learning* 109(2), pp. 373–440. doi: 10.1007/s10994-019-05855-6.

Esteva, A. et al. 2021. Deep learning-enabled medical computer vision. *npj Digital Medicine* 4(1), p. 5. doi: 10.1038/s41746-020-00376-2.

Esteva, A., Kuprel, B., Novoa, R.A., Ko, J., Swetter, S.M., Blau, H.M. and Thrun, S. 2017. Dermatologist-level classification of skin cancer with deep neural networks. *Nature* 542(7639), pp. 115–118. doi: 10.1038/nature21056.

FAO. 2020. *The State of World Fisheries and Aquaculture 2020*. FAO. Available at: http://www.fao.org/documents/card/en/c/ca9229en [Accessed: 28 May 2025].

Ferentinos, K.P. 2018. Deep learning models for plant disease detection and diagnosis. *Computers and Electronics in Agriculture* 145, pp. 311–318. doi: 10.1016/j.compag.2018.01.009.

Ford, R.B. and Mazzaferro, E.M. 2012. Clinical Signs. *Kirk & Bistner's Handbook of Veterinary Procedures and Emergency Treatment*, pp. 381–441. doi: 10.1016/B978-1-4377-0798-4.00003-7.

Fox, N. et al. 2020. "photosearcher" package in R: An accessible and reproducible method for harvesting large datasets from Flickr. *SoftwareX* 12, p. 100624. doi: 10.1016/j.softx.2020.100624.

Francis, M. and Deisy, C. 2019. Disease Detection and Classification in Agricultural Plants Using Convolutional Neural Networks — A Visual Understanding. pp. 1063–1068. doi: 10.1109/SPIN.2019.8711701.

Fuentes, A., Yoon, S., Kim, S.C. and Park, D.S. 2017. A robust deep-learning-based detector for real-time tomato plant diseases and pests recognition. *Sensors* 17(9), p. 2022.

Fuentes, S., Gonzalez Viejo, C., Tongson, E. and Dunshea, F.R. 2022. The livestock farming digital transformation: implementation of new and emerging technologies using artificial intelligence. *Animal Health Research Reviews* 23(1), pp. 59–71.

Gartenberg, C. 2018. *Flickr will end 1TB of free storage and limit free users to 1,000 photos*. Available at: https://www.theverge.com/2018/11/1/18051950/flickr-1000-photo-limit-free-accounts-changes-pro-subscription-smugmug [Accessed: 8 May 2025].

Gauthier, D.T. 2015. Bacterial zoonoses of fishes: A review and appraisal of evidence for linkages between fish and human infections. *The Veterinary Journal* 203(1), pp. 27–35. doi: 10.1016/j.tvjl.2014.10.028.

Gebru, T., Morgenstern, J., Vecchione, B., Vaughan, J.W., Wallach, H., III, H.D. and Crawford, K. 2021. Datasheets for datasets. *Commun. ACM* 64(12), pp. 86–92. doi: 10.1145/3458723.

Géron, A. 2022. *Hands-on machine learning with Scikit-Learn, Keras and TensorFlow: concepts, tools, and techniques to build intelligent systems*. Third edition. Sebastopol, California: O'Reilly Media, Inc.

Geurts, E.M., Reynolds, J.D. and Starzomski, B.M. 2023. Turning observations into biodiversity data: Broadscale spatial biases in community science. *Ecosphere* 14(6), p. e4582. doi: 10.1002/ecs2.4582.

Gómez-Villa, A., Salazar, A. and Vargas-Bonilla, J.F. 2016. Towards automatic wild animal monitoring: Identification of animal species in camera-trap images using very deep convolutional neural networks. *Ecol. Informatics* 41, pp. 24–32.

Green, S.E., Rees, J.P., Stephens, P.A., Hill, R.A. and Giordano, A.J. 2020. Innovations in Camera Trapping Technology and Approaches: The Integration of Citizen Science and Artificial Intelligence. *Animals* 10(1). doi: 10.3390/ani10010132.

Groom, Q., Weatherdon, L. and Geijzendorffer, I.R. 2017. Is citizen science an open science in the case of biodiversity observations? *Journal of Applied Ecology* 54(2), pp. 612–617. doi: 10.1111/1365-2664.12767.

Gui, P., Dang, W., Zhu, F. and Zhao, Q. 2021. Towards automatic field plant disease recognition. *Comput. Electron. Agric.* 191(C), p. 10.

Guo, C., Pleiss, G., Sun, Y. and Weinberger, K.Q. 2017. On Calibration of Modern Neural Networks. In: *Proceedings of the 34th International Conference on Machine Learning*. PMLR, pp. 1321–1330. Available at: https://proceedings.mlr.press/v70/guo17a.html [Accessed: 9 June 2025].

Gupta, A., Bringsdal, E., Knausgård, K.M. and Goodwin, M. 2022. Accurate Wound and Lice Detection in Atlantic Salmon Fish Using a Convolutional Neural Network. *Fishes* 7(6). doi: 10.3390/fishes7060345.

Haenssle, H.A. et al. 2018. Man against machine: diagnostic performance of a deep learning convolutional neural network for dermoscopic melanoma recognition in comparison to 58 dermatologists. *Annals of oncology* 29(8), pp. 1836–1842.

Haley, N.J. and Hoover, E.A. 2015. Chronic Wasting Disease of Cervids: Current Knowledge and Future Perspectives. *Annual Review of Animal Biosciences* 3(Volume 3, 2015), pp. 305–325. doi: 10.1146/annurev-animal-022114-111001.

Hardisty, A.R. et al. 2022. Digital Extended Specimens: Enabling an Extensible Network of Biodiversity Data Records as Integrated Digital Objects on the Internet. *BioScience* 72(10), pp. 978–987. doi: 10.1093/biosci/biac060.

Hasan, N., Ibrahim, S. and Aqilah Azlan, A. 2022. Fish diseases detection using convolutional neural network (CNN). *International Journal of Nonlinear Analysis and Applications* 13(1), pp. 1977–1984. doi: 10.22075/ijnaa.2022.5839.

Hawkins, C.E. et al. 2006. Emerging disease and population decline of an island endemic, the Tasmanian devil Sarcophilus harrisii. *Biological Conservation* 131(2), pp. 307–324. doi: https://doi.org/10.1016/j.biocon.2006.04.010.

He, K., Fan, H., Wu, Y., Xie, S. and Girshick, R. 2020. Momentum Contrast for Unsupervised Visual Representation Learning. Available at: http://arxiv.org/abs/1911.05722 [Accessed: 17 April 2025].

He, K., Gkioxari, G., Dollár, P. and Girshick, R. 2017. Mask R-CNN. p. arXiv:1703.06870.

He, K., Zhang, X., Ren, S. and Sun, J. 2016. Deep residual learning for image recognition. In: *Proceedings of the IEEE conference on computer vision and pattern recognition*. pp. 770–778. doi: 10.1109/CVPR.2016.90.

Heberling, J.M., Miller, J.T., Noesgaard, D., Weingart, S.B. and Schigel, D. 2021. Data integration enables global biodiversity synthesis. *Proceedings of the National Academy of Sciences* 118(6), p. e2018093118. doi: 10.1073/pnas.2018093118.

Hemelings, R. et al. 2020. Accurate prediction of glaucoma from colour fundus images with a convolutional neural network that relies on active and transfer learning. *Acta ophthalmologica* 98(1), pp. e94–e100.

Hemphill, L., Schöpke-Gonzalez, A. and Panda, A. 2022. Comparative sensitivity of social media data and their acceptable use in research. *Scientific Data* 9(1), p. 643. doi: 10.1038/s41597-022-01773-w.

HerdVision. 2021. *HerdVision | Automated cow health and productivity monitoring*. Available at: https://herd.vision/ [Accessed: 4 February 2026].

Hofstra, G. et al. 2023. Automated monitoring and detection of disease using a generic facial feature scoring system – A case study on FMD infected cows. *Preventive Veterinary Medicine* 213, p. 105880. doi: 10.1016/j.prevetmed.2023.105880.

Hospedales, T., Antoniou, A., Micaelli, P. and Storkey, A. 2022. Meta-Learning in Neural Networks: A Survey. *IEEE Transactions on Pattern Analysis & Machine Intelligence* 44(09), pp. 5149–5169. doi: 10.1109/TPAMI.2021.3079209.

Howard, A. et al. 2019. Searching for MobileNetV3. IEEE, pp. 1314–1324.

Hoyt, J.R., Kilpatrick, A.M. and Langwig, K.E. 2021. Ecology and impacts of white-nose syndrome on bats. *Nature Reviews Microbiology* 19(3), pp. 196–210. doi: 10.1038/s41579-020-00493-5.

Huang, Y.-P. and Khabusi, S.P. 2023. A CNN-OSELM Multi-Layer Fusion Network With Attention Mechanism for Fish Disease Recognition in Aquaculture. *IEEE Access* 11, pp. 58729–58744. doi: 10.1109/ACCESS.2023.3280540.

Hufschmid, J., Handasyde, K.A. and Beveridge, I. 2010. The role of host and environmental factors in the epidemiology of rumpwear in brushtail possums. *Australian Journal of Zoology* 58(4), pp. 250–262. doi: https://doi.org/10.1071/ZO10030.

IPTC Photo Metadata Working Group. 2024. *IPTC Photo Metadata Standard 2024.1*. Available at: https://iptc.org/std/photometadata/specification/IPTC-PhotoMetadata [Accessed: 9 July 2025].

Isaac, N.J.B. and Pocock, M.J.O. 2015. Bias and information in biological records. *Biological Journal of the Linnean Society* 115(3), pp. 522–531. doi: 10.1111/bij.12532.

Jagiello, Z., Dylewski, Ł. and Szulkin, M. 2024. The plastic homes of hermit crabs in the Anthropocene. *Science of The Total Environment* 913, p. 168959. doi: 10.1016/j.scitotenv.2023.168959.

Jain, R., Gupta, M., Taneja, S. and Hemanth, D.J. 2021. Deep learning based detection and analysis of COVID-19 on chest X-ray images. *Applied Intelligence* 51(3), pp. 1690–1700. doi: 10.1007/s10489-020-01902-1.

Jarić, I. et al. 2020. iEcology: Harnessing Large Online Resources to Generate Ecological Insights. *Trends in Ecology & Evolution* 35(7), pp. 630–639. doi: https://doi.org/10.1016/j.tree.2020.03.003.

Jones, K.E., Patel, N.G., Levy, M.A., Storeygard, A., Balk, D., Gittleman, J.L. and Daszak, P. 2008a. Global trends in emerging infectious diseases. *Nature* 451(7181), pp. 990–993. doi: 10.1038/nature06536.

Jones, M.E. et al. 2008b. Life-history change in disease-ravaged Tasmanian devil populations. *Proceedings of the National Academy of Sciences of the United States of America* 105(29), pp. 10023–10027. doi: 10.1073/pnas.0711236105.

Jorquera-Chavez, M., Fuentes, S., Dunshea, F.R., Warner, R.D., Poblete, T., Morrison, R.S. and Jongman, E.C. 2020. Remotely sensed imagery for early detection of respiratory disease in pigs: A pilot study. *Animals* 10(3). doi: 10.3390/ani10030451.

Karesh, W.B. et al. 2012. Ecology of zoonoses: natural and unnatural histories. *The Lancet* 380(9857), pp. 1936–1945. doi: 10.1016/S0140-6736(12)61678-X.

Khan, A., Sohail, A., Zahoora, U. and Qureshi, A.S. 2020. A survey of the recent architectures of deep convolutional neural networks. *Artificial Intelligence Review* 53(8), pp. 5455–5516. doi: 10.1007/s10462-020-09825-6.

Klemetsen, A., Amundsen, P.-A., Dempson, J.B., Jonsson, B., Jonsson, N., O'Connell, M.F. and Mortensen, E. 2003. Atlantic salmon Salmo salar L., brown trout Salmo trutta L. and Arctic charr Salvelinus alpinus (L.): a review of aspects of their life histories. *Ecology of Freshwater Fish* 12(1), pp. 1–59. doi: 10.1034/j.1600-0633.2003.00010.x.

Krizhevsky, A., Sutskever, I. and Hinton, G.E. 2012. Imagenet classification with deep convolutional neural networks. *Advances in neural information processing systems* 25, pp. 1097–1105.

Kucur, Ş.S., Holló, G. and Sznitman, R. 2018. A deep learning approach to automatic detection of early glaucoma from visual fields. *PLoS One* 13(11), p. e0206081.

Kumaar, A.S., Vignesh, A.V. and Deepak, K. 2024. FishNet Freshwater Fish Disease Detection using Deep Learning Techniques. In: *2024 2nd International Conference on Advancement in Computation & Computer Technologies (InCACCT)*. pp. 368–373. doi: 10.1109/InCACCT61598.2024.10550979.

Labelbox. 2025. Available at: https://labelbox.com.

Lawson, B., Petrovan, S.O. and Cunningham, A.A. 2015. Citizen Science and Wildlife Disease Surveillance. *EcoHealth* 12(4), pp. 693–702. doi: 10.1007/s10393-015-1054-z.

LeCun, Y., Bengio, Y. and Hinton, G. 2015. Deep learning. *Nature* 521(7553), pp. 436–444. doi: 10.1038/nature14539.

Lendemer, J. et al. 2020. The Extended Specimen Network: A Strategy to Enhance US Biodiversity Collections, Promote Research and Education. *BioScience* 70(1), pp. 23–30. doi: 10.1093/biosci/biz140.

Lennox, R.J. et al. 2022. Digital fisheries data in the Internet age: Emerging tools for research and monitoring using online data in recreational fisheries. *Fish and Fisheries* 23(4), pp. 926–940. doi: 10.1111/faf.12663.

Lenth, R.V. 2024. emmeans: Estimated Marginal Means, aka Least-Squares Means. Available at: https://CRAN.R-project.org/package=emmeans.

Li, D. et al. 2020. A Recognition Method for Rice Plant Diseases and Pests Video Detection Based on Deep Convolutional Neural Network. *Sensors* 20(3), p. 578.

Lin, T.-Y., Goyal, P., Girshick, R., He, K. and Dollár, P. 2020. Focal Loss for Dense Object Detection. *IEEE Transactions on Pattern Analysis and Machine Intelligence* 42(2), pp. 318–327. doi: 10.1109/TPAMI.2018.2858826.

Lindenmayer, D.B. and Likens, G.E. 2010. The science and application of ecological monitoring. *Biological Conservation* 143(6), pp. 1317–1328. doi: 10.1016/j.biocon.2010.02.013.

Liu, C. et al. 2023a. Research progress of computer vision technology in abnormal fish detection. *Aquacultural Engineering* 103, p. 102350. doi: 10.1016/j.aquaeng.2023.102350.

Liu, H., Li, C., Wu, Q. and Lee, Y.J. 2023b. Visual Instruction Tuning. *ArXiv* abs/2304.08485.

Liu, Y. et al. 2020. A deep learning system for differential diagnosis of skin diseases. *Nature Medicine* 26(6), pp. 900–908.

London School of Economics. 2025. *ethics-Using-internet-and-Social-media-data.pdf*. Available at: https://info.lse.ac.uk/staff/divisions/research-and-innovation/research/Assets/Documents/PDF/ethics-Using-internet-and-Social-media-data.pdf [Accessed: 9 July 2025].

Lüdecke, D., Ben-Shachar, M.S., Patil, I., Waggoner, P. and Makowski, D. 2021. performance: An R package for assessment, comparison and testing of statistical models. *Journal of open source software* 6(60).

Luo, C.-Y., Pearson, P., Xu, G. and Rich, S.M. 2022. A Computer Vision-Based Approach for Tick Identification Using Deep Learning Models. *Insects* 13(2), p. 116. doi: 10.3390/insects13020116.

Lürig, M.D., Donoughe, S., Svensson, E.I., Porto, A. and Tsuboi, M. 2021. Computer Vision, Machine Learning, and the Promise of Phenomics in Ecology and Evolutionary Biology. *Frontiers in Ecology and Evolution* 9(148). doi: 10.3389/fevo.2021.642774.

Lynch, A.J. et al. 2016. The social, economic, and environmental importance of inland fish and fisheries. *Environmental Reviews* 24(2), pp. 115–121. doi: 10.1139/er-2015-0064.

van der Maaten, L.J.P. and Hinton, G.E. 2008. Visualizing High-Dimensional Data Using t-SNE. *Journal of Machine Learning Research* 9(nov), pp. 2579–2605.

MacAulay, S., Ellison, A.R., Kille, P. and Cable, J. 2022. Moving towards improved surveillance and earlier diagnosis of aquatic pathogens: From traditional methods to emerging technologies. *Reviews in Aquaculture* 14(4), pp. 1813–1829. doi: https://doi.org/10.1111/raq.12674.

Malik, S., Kumar, T. and Sahoo, A.K. 2017. Image processing techniques for identification of fish disease. In: *2017 IEEE 2nd International Conference on Signal and Image Processing, ICSIP 2017*. Institute of Electrical and Electronics Engineers Inc., pp. 55–59. doi: 10.1109/SIPROCESS.2017.8124505.

Marshall, B.M. et al. 2020. An inventory of online reptile images. *Zootaxa* 4896(2), p. zootaxa.4896.2.6. doi: 10.11646/zootaxa.4896.2.6.

Maruf, A.A., Fahim, S.H., Bashar, R., Rumy, R.A., Chowdhury, S.I. and Aung, Z. 2024. Classification of Freshwater Fish Diseases in Bangladesh Using a Novel Ensemble Deep Learning Model: Enhancing Accuracy and Interpretability. *IEEE Access* 12, pp. 96411–96435. doi: 10.1109/ACCESS.2024.3426041.

Matthews, E. 2019. Saprolegnia *in wild fish*. PhD thesis, Cardiff University.

Matthews, E., Ellison, A. and Cable, J. 2021. *Saprolegnia parasitica* zoospore activity and host survival indicates isolate variation in host preference. *Fungal Biology* 125(4), pp. 260–268. doi: https://doi.org/10.1016/j.funbio.2020.11.002.

McCallum, H. 2008. Tasmanian devil facial tumour disease: lessons for conservation biology. *Trends in Ecology & Evolution* 23(11), pp. 631–637. doi: 10.1016/j.tree.2008.07.001.

Mia, Md.J., Mahmud, R.B., Sadad, Md.S., Asad, H.A. and Hossain, R. 2022. An in-depth automated approach for fish disease recognition. *Journal of King Saud University - Computer and Information Sciences* 34(9), pp. 7174–7183. doi: 10.1016/j.jksuci.2022.02.023.

Miao, Z. et al. 2019. Insights and approaches using deep learning to classify wildlife. *Scientific Reports* 9(1), p. 8137. doi: 10.1038/s41598-019-44565-w.

Mittal, A., Moorthy, A.K. and Bovik, A.C. 2012. No-Reference Image Quality Assessment in the Spatial Domain. *IEEE Transactions on Image Processing* 21(12), pp. 4695–4708. doi: 10.1109/TIP.2012.2214050.

Mittal, A., Soundararajan, R. and Bovik, A.C. 2013. Making a “Completely Blind” Image Quality Analyzer. *IEEE Signal Processing Letters* 20(3), pp. 209–212. doi: 10.1109/LSP.2012.2227726.

Mohanty, S.P., Hughes, D.P. and Salathé, M. 2016. Using Deep Learning for Image-Based Plant Disease Detection. *Frontiers in Plant Science* 7(1419). doi: 10.3389/fpls.2016.01419.

Monkman, G.G., Kaiser, M.J. and Hyder, K. 2018. Text and data mining of social media to map wildlife recreation activity. *Biological Conservation* 228, pp. 89–99. doi: 10.1016/j.biocon.2018.10.010.

Morcatty, T.Q. et al. 2024. Navigating ethical challenges in online wildlife trade research. *Conservation Biology* 38(5), p. e14341. doi: 10.1111/cobi.14341.

Morens, D.M., Folkers, G.K. and Fauci, A.S. 2004. The challenge of emerging and re-emerging infectious diseases. *Nature* 430(6996), pp. 242–249. doi: 10.1038/nature02759.

Mörner, T., Obendorf, D.L., Artois, M. and Woodford, M.H. 2002. Surveillance and monitoring of wildlife diseases. *Revue Scientifique Et Technique (International Office of Epizootics)* 21(1), pp. 67–76. doi: 10.20506/rst.21.1.1321.

Mumuni, A. and Mumuni, F. 2022. Data augmentation: A comprehensive survey of modern approaches. *Array* 16, p. 100258. doi: 10.1016/j.array.2022.100258.

Muneza, A.B. et al. 2019. QUANTIFYING THE SEVERITY OF GIRAFFE SKIN DISEASE VIA PHOTOGRAMMETRY ANALYSIS OF CAMERA TRAP DATA. *Journal of Wildlife Diseases* 55(4), pp. 770–781. doi: 10.7589/2018-06-149.

Murphy, C.J., Collier, M.A., Jacoby, A.-M., Patterson, E.M., Wallen, M.M., Mann, J. and Bansal, S. 2025. Automated skin lesion detection and prevalence estimation in Tamanend's bottlenose dolphins. *Ecological Informatics* 85, p. 102942. doi: 10.1016/j.ecoinf.2024.102942.

Murray, M.H., Fidino, M., Lehrer, E.W., Simonis, J.L. and Magle, S.B. 2021. A multi-state occupancy model to non-invasively monitor visible signs of wildlife health with camera traps that accounts for image quality. *Journal of Animal Ecology* 90(8), pp. 1973–1984. doi: https://doi.org/10.1111/1365-2656.13515.

Myint, B.B., Onizuka, T., Tin, P., Aikawa, M., Kobayashi, I. and Zin, T.T. 2024. Development of a real-time cattle lameness detection system using a single side-view camera. *Scientific Reports* 14(1), p. 13734. doi: 10.1038/s41598-024-64664-7.

Naiman, R.J., Bilby, R.E., Schindler, D.E. and Helfield, J.M. 2002. Pacific Salmon, Nutrients, and the Dynamics of Freshwater and Riparian Ecosystems. *Ecosystems* 5(4), pp. 399–417. doi: 10.1007/s10021-001-0083-3.

Nam, M.G. and Dong, S.-Y. 2023. Classification of Companion Animals' Ocular Diseases: Domain Adversarial Learning for Imbalanced Data. *IEEE Access* 11, pp. 143948–143955. doi: 10.1109/ACCESS.2023.3344579.

Nazir, S. and Kaleem, M. 2021. Advances in image acquisition and processing technologies transforming animal ecological studies. *Ecological Informatics* 61, p. 101212. doi: 10.1016/j.ecoinf.2021.101212.

Nemoto, K., Hamaguchi, R., Imaizumi, T. and Hikosaka, S. 2018. Classification of rare building change using CNN with multi-class focal loss. In: *IGARSS 2018 - 2018 IEEE international geoscience and remote sensing symposium*. pp. 4663–4666. doi: 10.1109/IGARSS.2018.8517563.

Nie, Y., Sommella, P., Carratù, M., O'Nils, M. and Lundgren, J. 2023. A deep CNN transformer hybrid model for skin lesion classification of dermoscopic images using focal loss. *Diagnostics* 13(1). Available at: https://www.mdpi.com/2075-4418/13/1/72.

Nishio, M., Nishizawa, M., Sugiyama, O., Kojima, R., Yakami, M., Kuroda, T. and Togashi, K. 2018. Computer-aided diagnosis of lung nodule using gradient tree boosting and Bayesian optimization. *PLoS One* 13(4), p. e0195875. doi: 10.1371/journal.pone.0195875.

Niu, Z., Zhou, M., Wang, L., Gao, X. and Hua, G. 2016. Ordinal Regression with Multiple Output CNN for Age Estimation. In: *2016 IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*. Las Vegas, NV, USA: IEEE, pp. 4920–4928. Available at: http://ieeexplore.ieee.org/document/7780901/ [Accessed: 9 June 2025].

Noga, E.J. 2010. *Fish Disease: Diagnosis and Treatment*. Newark, UNITED STATES: John Wiley & Sons, Incorporated. Available at: http://ebookcentral.proquest.com/lib/cardiff/detail.action?docID=700344 [Accessed: 14 April 2025].

Norouzzadeh, M.S., Morris, D., Beery, S., Joshi, N., Jojic, N. and Clune, J. 2021. A deep active learning system for species identification and counting in camera trap images. *Methods in Ecology and Evolution* 12(1), pp. 150–161. doi: https://doi.org/10.1111/2041-210X.13504.

Norouzzadeh, M.S., Nguyen, A., Kosmala, M., Swanson, A., Palmer, M.S., Packer, C. and Clune, J. 2018. Automatically identifying, counting, and describing wild animals in camera-trap images with deep learning. *Proceedings of the National Academy of Sciences* 115(25), p. E5716. doi: 10.1073/pnas.1719367115.

Nunes, J.A.C.C., Cruz, I.C.S., Nunes, A. and Pinheiro, H.T. 2020. Speeding up coral reef conservation with AI-aided automated image analysis. *Nature Machine Intelligence* 2(6), pp. 292–292. doi: 10.1038/s42256-020-0192-3.

Nurçin, F.V., Sentürk, N., Imanov, E., Thalmann, S. and Fagg, K. 2024. Automated Tasmanian devil segmentation and devil facial tumour disease classification. *Wildlife Research* 51(1). Available at: https://doi.org/10.1071/WR22155.

Olsen, A.S., Cook, N. and Perkins, S.E. 2025. SalmoSapro Metadata Catalog: A Cross-Platform Index of Salmonid Images with *Saprolegnia* Classifications. Available at: https://zenodo.org/uploads/15097673.

Olsen, A.S., Rosin, P.L., Jones, C.B., Cable, J. and Perkins, S.E. 2026. Computer vision for infectious disease surveillance; *Saprolegnia* spp. in salmonids. *Ecological Informatics* 93, p. 103567. doi: 10.1016/j.ecoinf.2025.103567.

O'Mahony, N. et al. 2020. Deep Learning vs. Traditional Computer Vision. In: Arai, K. and Kapoor, S. eds *Advances in Computer Vision*. Advances in Intelligent Systems and Computing. Cham: Springer International Publishing, pp. 128–144. doi: 10.1007/978-3-030-17795-9_10.

Palmer, M.S., Huebner, S.E., Willi, M., Fortson, L. and Packer, C. 2021. Crowdsourcing, computing, and conservation: how citizen science and artificial intelligence can improve the use of camera trap data to tackle large-scale ecological challenges.

Panda, R., Puhan, N.B., Rao, A., Mandal, B., Padhy, D. and Panda, G. 2018. Deep convolutional neural network-based patch classification for retinal nerve fiber layer defect detection in early glaucoma. *Journal of Medical Imaging* 5(4), p. 044003.

Pantazis, O. et al. 2024. Deep learning-based ecological analysis of camera trap images is impacted by training data quality and size. *ArXiv* abs/2408.14348.

Park, J., Oh, M. and Han, S. 2007. Fish Disease Diagnosis System Based on Image Processing of Pathogens' Microscopic Images. In: *2007 Frontiers in the Convergence of Bioscience and Information Technologies*. pp. 878–883. doi: 10.1109/FBIT.2007.157.

Pavić, D. et al. 2022. Tracing the oomycete pathogen *Saprolegnia parasitica* in aquaculture and the environment. *Scientific Reports* 12(1), p. 16646. doi: 10.1038/s41598-022-16553-0.

Pech-Pacheco, J.L., Cristobal, G., Chamorro-Martinez, J. and Fernandez-Valdivia, J. 2000. Diatom autofocusing in brightfield microscopy: a comparative study. In: *Proceedings 15th International Conference on Pattern Recognition. ICPR-2000*. pp. 314–317 vol.3. Available at: https://ieeexplore.ieee.org/document/903548 [Accessed: 10 June 2025].

Pecl, G.T. et al. 2017. Biodiversity redistribution under climate change: Impacts on ecosystems and human well-being. *Science* 355(6332), p. eaai9214. doi: 10.1126/science.aai9214.

Peeler, E.J., Oidtmann, B.C., Midtlyng, P.J., Miossec, L. and Gozlan, R.E. 2011. Non-native aquatic animals introductions have driven disease emergence in Europe. *Biological Invasions* 13(6), pp. 1291–1303. doi: 10.1007/s10530-010-9890-9.

Pernat, N., Memedemin, D., August, T., Preda, C., Reyserhove, L., Schirmel, J. and Groom, Q. 2024. Extracting secondary data from citizen science images reveals host flower preferences of the Mexican grass-carrying wasp Isodontia mexicana in its native and introduced ranges. *Ecology and Evolution* 14(6), p. e11537. doi: 10.1002/ece3.11537.

Petmezas, G., Cheimariotis, G.-A., Stefanopoulos, L., Rocha, B., Paiva, R.P., Katsaggelos, A.K. and Maglaveras, N. 2022. Automated lung sound classification using a hybrid CNN-LSTM network and focal loss function. *Sensors* 22(3). Available at: https://www.mdpi.com/1424-8220/22/3/1232.

Phelps, K.L. et al. 2019. Bat Research Networks and Viral Surveillance: Gaps and Opportunities in Western Asia. *Viruses* 11(3). doi: 10.3390/v11030240.

Pokharel, S.S., Sharma, N. and Sukumar, R. 2022. Viewing the rare through public lenses: insights into dead calf carrying and other thanatological responses in Asian elephants using YouTube videos. *Royal Society Open Science* 9(5), p. 211740. doi: 10.1098/rsos.211740.

Pollock, L.J. et al. 2025. Harnessing artificial intelligence to fill global shortfalls in biodiversity knowledge. *Nature Reviews Biodiversity* 1(3), pp. 166–182. doi: 10.1038/s44358-025-00022-3.

Poulin, R. et al. 2021. iParasitology: Mining the Internet to Test Parasitological Hypotheses. *Trends in Parasitology* 37(4), pp. 267–272. doi: https://doi.org/10.1016/j.pt.2021.01.003.

Pyšek, P. et al. 2020. Scientists' warning on invasive alien species. *Biological Reviews* 95(6), pp. 1511–1534. doi: 10.1111/brv.12627.

R Core Team. 2021. R: A Language and Environment for Statistical Computing. Available at: https://www.R-project.org.

Rachman, F., Akbar, M.N.S. and Putera, E. 2023. Fish Disease Detection of Epizootic Ulcerative Syndrome Using Deep Learning Image Processing Technique. *Proceedings*

International Conference on Fisheries and Aquaculture 8(1), pp. 23–34. doi: 10.17501/23861282.2023.8102.

Radford, A. et al. 2021. Learning Transferable Visual Models From Natural Language Supervision. In: *International Conference on Machine Learning*.

Rafiq, K., Beery, S., Palmer, M.S., Harchaoui, Z. and Abrahms, B. 2025. Generative AI as a tool to accelerate the field of ecology. *Nature Ecology & Evolution* 9(3), pp. 378–385. doi: 10.1038/s41559-024-02623-1.

Rajkomar, A., Lingam, S., Taylor, A.G., Blum, M. and Mongan, J. 2017. High-Throughput Classification of Radiographs Using Deep Convolutional Neural Networks. *Journal of Digital Imaging* 30(1), pp. 95–101. doi: 10.1007/s10278-016-9914-9.

Rana, M. and Bhushan, M. 2023. Machine learning and deep learning approach for medical image analysis: diagnosis to detection. *Multimedia Tools and Applications* 82(17), pp. 26731–26769. doi: 10.1007/s11042-022-14305-w.

Ratnayake, M.N., Dyer, A.G. and Dorin, A. 2021. Tracking individual honeybees among wildflower clusters with computer vision-facilitated pollinator monitoring. *PLoS One* 16(2), pp. e0239504–e0239504. doi: 10.1371/journal.pone.0239504.

Ravoor, P.C. and Sudarshan, T.S.B. 2020. Deep Learning Methods for Multi-Species Animal Re-identification and Tracking – a Survey. *Computer Science Review* 38, p. 100289. doi: https://doi.org/10.1016/j.cosrev.2020.100289.

Redmon, J., Divvala, S., Girshick, R. and Farhadi, A. 2015. You Only Look Once: Unified, Real-Time Object Detection. p. arXiv:1506.02640.

Reid, A.J. et al. 2019. Emerging threats and persistent conservation challenges for freshwater biodiversity. *Biological Reviews* 94(3), pp. 849–873. doi: 10.1111/brv.12480.

Reimers, N. and Gurevych, I. 2019. Sentence-BERT: Sentence Embeddings using Siamese BERT-Networks. Available at: http://arxiv.org/abs/1908.10084 [Accessed: 4 June 2025].

Ren, S., He, K., Girshick, R. and Sun, J. 2015. Faster R-CNN: Towards Real-Time Object Detection with Region Proposal Networks. p. arXiv:1506.01497.

Reyserhove, L. et al. 2020. A checklist recipe: making species data open and FAIR. *Database* 2020, p. baaa084. doi: 10.1093/database/baaa084.

Ringwaldt, E.M., Brook, B.W., Carver, S. and Buettel, J.C. 2022. Signs of Rumpwear in the Common Ringtail Possum, Pseudocheirus peregrinus. *Journal of Wildlife Diseases* 58(3), pp. 675–679. doi: 10.7589/jwd-d-21-00115.

Ringwaldt, E.M., Buettel, J.C., Carver, S. and Brook, B.W. 2025. Epidemiological Dynamics of a Visually Apparent Disease: Camera Trapping and Machine-Learning Applied to Rumpwear in the Common Brushtail Possum. *Integrative Zoology* n/a(n/a). Available at: https://onlinelibrary.wiley.com/doi/abs/10.1111/1749-4877.12995 [Accessed: 3 June 2025].

Robie, A.A., Seagraves, K.M., Egnor, S.E.R. and Branson, K. 2017. Machine vision methods for analyzing social interactions. *The Journal of Experimental Biology* 220(1), p. 25. doi: 10.1242/jeb.142281.

Robinson, R.A. et al. 2010. Emerging Infectious Disease Leads to Rapid Population Declines of Common British Birds. *PLOS ONE* 5(8), p. e12215. doi: 10.1371/journal.pone.0012215.

Rony, M., Barai, D., Riad and Hasan, Z. 2021. Cattle External Disease Classification Using Deep Learning Techniques. In: *2021 12th International Conference on Computing Communication and Networking Technologies (ICCCNT)*. pp. 1–7. doi: 10.1109/ICCCNT51525.2021.9579662.

Rousseeuw, P.J. 1987. Silhouettes: A graphical aid to the interpretation and validation of cluster analysis. *Journal of Computational and Applied Mathematics* 20, pp. 53–65. doi: 10.1016/0377-0427(87)90125-7.

Savary, S., Willocquet, L., Pethybridge, S.J., Esker, P., McRoberts, N. and Nelson, A. 2019. The global burden of pathogens and pests on major food crops. *Nature Ecology & Evolution* 3(3), pp. 430–439. doi: 10.1038/s41559-018-0793-y.

Schilling, A.-K., Mazzamuto, M.V. and Romeo, C. 2022. A Review of Non-Invasive Sampling in Wildlife Disease and Health Research: What's New? *Animals* 12(13), p. 1719. doi: 10.3390/ani12131719.

Schindler, D.E., Scheuerell, M.D., Moore, J.W., Gende, S.M., Francis, T.B. and Palen, W.J. 2003. Pacific salmon and the ecology of coastal ecosystems. *Frontiers in Ecology and the Environment* 1(1), pp. 31–37. doi: 10.1890/1540-9295(2003)001%5B0031:PSATEO%5D2.0.CO;2.

Schneider, S., Greenberg, S., Taylor, G.W. and Kremer, S.C. 2020. Three critical factors affecting automated image species recognition performance for camera traps. *Ecology and Evolution* 10(7), pp. 3503–3517. doi: https://doi.org/10.1002/ece3.6147.

Schneider, S., Taylor, G.W., Linquist, S. and Kremer, S.C. 2019. Past, present and future approaches using computer vision for animal re-identification from camera trap data. *Methods in Ecology and Evolution* 10(4), pp. 461–470. doi: https://doi.org/10.1111/2041-210X.13133.

Scott, D.M., Baker, R., Tomlinson, A., Berg, M.J., Charman, N. and Tolhurst, B.A. 2020. Spatial distribution of sarcoptic mange (*Sarcoptes scabiei*) in urban foxes (*Vulpes vulpes*) in Great Britain as determined by citizen science. *Urban Ecosystems* 23(5), pp. 1127–1140. doi: 10.1007/s11252-020-00985-5.

Selvaraju, R.R., Cogswell, M., Das, A., Vedantam, R., Parikh, D. and Batra, D. 2020. Grad-CAM: Visual Explanations from Deep Networks via Gradient-based Localization. *International Journal of Computer Vision* 128(2), pp. 336–359. doi: 10.1007/s11263-019-01228-7.

Shinn, A.P., Pratoomyot, J., Bron, J.E., Paladini, G., Brooker, E.E. and Brooker, A.J. 2014. Economic costs of protistan and metazoan parasites to global mariculture. *Parasitology* 142(1), pp. 196–270. doi: 10.1017/s0031182014001437.

Shorten, C. and Khoshgoftaar, T.M. 2019. A survey on Image Data Augmentation for Deep Learning. *Journal of Big Data* 6(1), p. 60. doi: 10.1186/s40537-019-0197-0.

Siachos, N., Lennox, M., Anagnostopoulos, A., Griffiths, B.E., Neary, J.M., Smith, R.F. and Oikonomou, G. 2024. Development and validation of a fully automated 2-dimensional imaging system generating body condition scores for dairy cows using machine learning. *Journal of Dairy Science* 107(4), pp. 2499–2511. doi: 10.3168/jds.2023-23894.

Simonyan, K. and Zisserman, A. 2014. Very deep convolutional networks for large-scale image recognition. *arXiv preprint arXiv:1409.1556*.

Singh, R., Bharti, V., Purohit, V., Kumar, A., Singh, A.K. and Singh, S.K. 2021. MetaMed: Few-shot medical image classification using gradient-based meta-learning. *Pattern Recognition* 120, p. 108111. doi: 10.1016/j.patcog.2021.108111.

Skirrow, R. 2024. *Fish on Film: the challenges of using AI to improve sustainable fisheries management – Marine Science*. Available at: https://marinescience.blog.gov.uk/2024/07/25/using-ai-to-improve-sustainable-fisheries-management/ [Accessed: 9 January 2026].

Skov, C. et al. 2021. Expert opinion on using angler Smartphone apps to inform marine fisheries management: status, prospects, and needs. *ICES Journal of Marine Science* 78(3), pp. 967–978. doi: 10.1093/icesjms/fsaa243.

Sladojevic, S., Arsenovic, M., Anderla, A., Culibrk, D. and Stefanovic, D. 2016a. Deep Neural Networks Based Recognition of Plant Diseases by Leaf Image Classification. Van Hulle, M. ed. *Computational Intelligence and Neuroscience* 2016, p. 3289801. doi: 10.1155/2016/3289801.

Sladojevic, S., Arsenovic, M., Anderla, A., Culibrk, D. and Stefanovic, D. 2016b. Deep Neural Networks Based Recognition of Plant Diseases by Leaf Image Classification. Van Hulle, M. ed. *Computational Intelligence and Neuroscience* 2016, p. 3289801. doi: 10.1155/2016/3289801.

Smith, A. et al. 2024. Computer vision model for the detection of canine pododermatitis and neoplasia of the paw. *Veterinary Dermatology* 35(2), pp. 138–147. doi: 10.1111/vde.13221.

Smith, K.F., Acevedo-Whitehouse, K. and Pedersen, A.B. 2009. The role of infectious diseases in biological conservation. *Animal Conservation* 12(1), pp. 1–12. doi: https://doi.org/10.1111/j.1469-1795.2008.00228.x.

Sohn, K. et al. 2020. FixMatch: Simplifying Semi-Supervised Learning with Consistency and Confidence. In: *Advances in Neural Information Processing Systems*. Curran Associates, Inc., pp. 596–608. Available at:

https://papers.nips.cc/paper/2020/hash/06964dce9addb1c5cb5d6e3d9838f733-Abstract.html [Accessed: 6 June 2025].

Speed, J.D.M., Bendiksby, M., Finstad, A.G., Hassel, K., Kolstad, A.L. and Prestø, T. 2018. Contrasting spatial, temporal and environmental patterns in observation and specimen based species occurrence data. *PLOS ONE* 13(4), p. e0196417. doi: 10.1371/journal.pone.0196417.

Spiesman, B.J. et al. 2021. Assessing the potential for deep learning and computer vision to identify bumble bee species from images. *Scientific Reports* 11(1), p. 7580. doi: 10.1038/s41598-021-87210-1.

Stallknecht, D.E. 2007. Impediments to Wildlife Disease Surveillance, Research, and Diagnostics. In: Childs, J. E., Mackenzie, J. S., and Richt, J. A. eds *Wildlife and Emerging Zoonotic Diseases: The Biology, Circumstances and Consequences of Cross-Species Transmission*. Berlin, Heidelberg: Springer, pp. 445–461. Available at: https://doi.org/10.1007/978-3-540-70962-6_17 [Accessed: 15 July 2025].

Steidl, M. 2016. *Many Social Media Sites Still Remove Image Rights Information From Photos*. Available at: https://iptc.org/news/many-social-media-sites-still-remove-image-rights-information-from-photos/ [Accessed: 9 July 2025].

Stephen, C. and Karesh, W.B. 2014. Is One Health delivering results? Introduction. *Revue scientifique et technique (International Office of Epizootics)* 33(2), pp. 375–392. doi: 10.20506/rst.33.2.2301.

Sun, Z. et al. 2021. Webly Supervised Fine-Grained Recognition: Benchmark Datasets and An Approach. In: *2021 IEEE/CVF International Conference on Computer Vision (ICCV)*. Montreal, QC, Canada: IEEE, pp. 10582–10591. Available at: https://ieeexplore.ieee.org/document/9710924/ [Accessed: 9 July 2025].

Swanson, A., Kosmala, M., Lintott, C., Simpson, R., Smith, A. and Packer, C. 2015. Snapshot Serengeti, high-frequency annotated camera trap images of 40 mammalian species in an African savanna. *Scientific Data* 2(1), p. 150026. doi: 10.1038/sdata.2015.26.

Szegedy, C., Vanhoucke, V., Ioffe, S., Shlens, J. and Wojna, Z. 2016. Rethinking the inception architecture for computer vision. pp. 2818–2826.

Szeliski, R. 2022. *Computer Vision: Algorithms and Applications*. Cham: Springer International Publishing. Available at: https://link.springer.com/10.1007/978-3-030-34372-9 [Accessed: 17 April 2025].

Szentivanyi, T. and Vincze, O. 2022. Tracking wildlife diseases using community science: an example through toad myiasis. *European Journal of Wildlife Research* 68(6), p. 74. doi: 10.1007/s10344-022-01623-5.

Tabak, M.A. et al. 2019. Machine learning to classify animal species in camera trap images: Applications in ecology. *Methods in Ecology and Evolution* 10(4), pp. 585–590. doi: https://doi.org/10.1111/2041-210X.13120.

Tan, M. and Le, Q. 2019. Efficientnet: Rethinking model scaling for convolutional neural networks. PMLR, pp. 6105–6114.

Tan, M. and Le, Q.V. 2021. EfficientNetV2: Smaller Models and Faster Training. Available at: https://arxiv.org/abs/2104.00298.

Tandel, R.S., Dash, P., Bhat, R.A.H., Sharma, P.C., Kalingapuram, K., Dubey, M. and Sarma, D. 2021. Morphological and molecular characterization of *Saprolegnia* spp. from Himalayan snow trout, *Schizothorax richardsonii*: A case study report. *Aquaculture (Amsterdam, Netherlands)* 531, p. 735824. doi: 10.1016/j.aquaculture.2020.735824.

Taylor, L.H., Latham, S.M. and Woolhouse, M.E. 2001. Risk factors for human disease emergence. *Philosophical Transactions of the Royal Society B: Biological Sciences* 356(1411), p. 983. doi: 10.1098/rstb.2001.0888.

Theobald, E.J. et al. 2015. Global change and local solutions: Tapping the unrealized potential of citizen science for biodiversity research. *Biological Conservation* 181, pp. 236–244. doi: 10.1016/j.biocon.2014.10.021.

Thorstad, E.B. et al. 2021. Atlantic salmon in a rapidly changing environment—Facing the challenges of reduced marine survival and climate change. *Aquatic Conservation: Marine and Freshwater Ecosystems* 31(9), pp. 2654–2665. doi: 10.1002/aqc.3624.

Tidal. 2025. *Product Tidal*. Available at: https://www.tidalx.ai/en/product [Accessed: 9 January 2026].

Toivonen, T. et al. 2019. Social media data for conservation science: A methodological overview. *Biological Conservation* 233, pp. 298–315. doi: 10.1016/j.biocon.2019.01.023.

Torney, C.J. et al. 2019. A comparison of deep learning and citizen science techniques for counting wildlife in aerial survey images. *Methods in Ecology and Evolution* 10(6), pp. 779–787. doi: https://doi.org/10.1111/2041-210X.13165.

Troudet, J., Grandcolas, P., Blin, A., Vignes-Lebbe, R. and Legendre, F. 2017. Taxonomic bias in biodiversity data and societal preferences. *Scientific Reports* 7(1), p. 9132. doi: 10.1038/s41598-017-09084-6.

Tuia, D. et al. 2022. Perspectives in machine learning for wildlife conservation. *Nature Communications* 13(1), p. 792. doi: 10.1038/s41467-022-27980-y.

Tulloch, A.I.T. et al. 2018. A decision tree for assessing the risks and benefits of publishing biodiversity data. *Nature Ecology & Evolution* 2(8), pp. 1209–1217. doi: 10.1038/s41559-018-0608-1.

Valletta, J.J., Torney, C., Kings, M., Thornton, A. and Madden, J. 2017. Applications of machine learning in animal behaviour studies. *Animal Behaviour* 124, pp. 203–220. doi: https://doi.org/10.1016/j.anbehav.2016.12.005.

Van Horn, G. et al. 2018. The iNaturalist Species Classification and Detection Dataset. *arXiv:1707.06642 [cs]*.

Vardi, R. et al. 2024. Leveraging social media and other online data to study animal behavior. *PLOS Biology* 22(8), p. e3002793. doi: 10.1371/journal.pbio.3002793.

Vaughan, P.M., Buettel, J.C. and Brook, B.W. 2022. Investigating Avian Behaviour Using Opportunistic Camera-Trap Imagery Reveals an Untapped Data Source. *Ornithological Science* 21(1). Available at: https://bioone.org/journals/ornithological-science/volume-21/issue-1/osj.21.3/Investigating-Avian-Behaviour-Using-Opportunistic-Camera-Trap-Imagery-Reveals-an/10.2326/osj.21.3.full [Accessed: 17 January 2024].

Vendrow, E. et al. 2024. INQUIRE: A Natural World Text-to-Image Retrieval Benchmark. *ArXiv* abs/2411.02537.

Vidal, M., Wolf, N., Rosenberg, B., Harris, B.P. and Mathis, A. 2021. Perspectives on Individual Animal Identification from Biology and Computer Vision. *Integrative and Comparative Biology* 61(3), pp. 900–916. doi: 10.1093/icb/icab107.

Vieira da Silva do Nascimento, I.T., Silva, J. dos S., Melo, T.A. de, Santos, D.M.S. and Serra, I.M.R. de S. 2020. Registry of saprolegniose in fish cultivated in the world: a compilation of data. *Research, Society and Development* 9(11), p. e959119556. doi: 10.33448/rsd-v9i11.9556.

Vijayalakshmi, M., Sasithradevi, A. and Prakash, P. 2023. Transfer Learning Approach for Epizootic Ulcerative Syndrome and *Ichthyophthirius* Disease Classification in Fish Species. In: *2023 International Conference on Bio Signals, Images, and Instrumentation (ICBSII)*. pp. 1–5. doi: 10.1109/ICBSII58188.2023.10181046.

Walmsley, T., Rose, A. and Wei, D. 2021. The Impacts of the Coronavirus on the Economy of the United States. *Economics of Disasters and Climate Change* 5(1), pp. 1–52. doi: 10.1007/s41885-020-00080-1.

Watsa, M. 2020. Rigorous wildlife disease surveillance. *Science* 369(6500), pp. 145–147. doi: doi:10.1126/science.abc0017.

Webster, M.S. 2017. *The Extended Specimen: Emerging Frontiers in Collections-Based Ornithological Research*. CRC Press.

Wehbe, R.M. et al. 2023. Deep Learning for Cardiovascular Imaging: A Review. *JAMA Cardiology* 8(11), pp. 1089–1098. doi: 10.1001/jamacardio.2023.3142.

Wei, X.-S. et al. 2022. Fine-Grained Image Analysis With Deep Learning: A Survey. *IEEE Transactions on Pattern Analysis and Machine Intelligence* 44(12), pp. 8927–8948. doi: 10.1109/TPAMI.2021.3126648.

Weinstein, B.G. 2018. A computer vision for animal ecology. *Journal of Animal Ecology* 87(3), pp. 533–545. doi: https://doi.org/10.1111/1365-2656.12780.

van West, P. and Beakes, G.W. 2014. Animal pathogenic oomycetes. *Fungal Biology* 118(7), pp. 525–526. doi: https://doi.org/10.1016/j.funbio.2014.05.004.

Weston, A.D. et al. 2019. Automated abdominal segmentation of CT scans for body composition analysis using deep learning. *Radiology* 290(3), pp. 669–679.

Wilkinson, M.D. et al. 2016. The FAIR Guiding Principles for scientific data management and stewardship. *Scientific Data* 3(1), p. 160018. doi: 10.1038/sdata.2016.18.

Willi, M. et al. 2019. Identifying animal species in camera trap images using deep learning and citizen science. *Methods in Ecology and Evolution* 10(1), pp. 80–91. doi: https://doi.org/10.1111/2041-210X.13099.

de Wit, C.D.G. and Johnson, P.T.J. 2024. Black Spot Syndrome in ocean surgeonfish: using video-based surveillance to quantify disease severity and test environmental drivers. *Marine Biology* 171(5), p. 110. doi: 10.1007/s00227-024-04426-1.

WOAH. 2020. *Animal Diseases*. Available at: https://www.woah.org/en/what-we-do/animal-health-and-welfare/animal-diseases/ [Accessed: 23 June 2025].

WOAH. 2023. *Terrestrial Animal Health Code - Recommendations applicable to WOAH listed diseases*. 31st edn. Available at: https://doc.woah.org/dyn/portal/index.xhtml?page=alo&aloId=43490 [Accessed: 12 July 2025].

WOAH and IUCN. 2024. *General Guidelines for Surveillance of Diseases, Pathogens and Toxic Agents in Free-ranging Wildlife*. World Organisation for Animal Health and IUCN. Available at: https://www.woah.org/app/uploads/2024/09/2024-final-guidelines-disease-pathogen-toxin-surv-wildlife-v27.06.pdf [Accessed: 12 July 2025].

Wu, D., Han, M., Song, H., Song, L. and Duan, Y. 2023. Monitoring the respiratory behavior of multiple cows based on computer vision and deep learning. *Journal of Dairy Science* 106(4), pp. 2963–2979. doi: 10.3168/jds.2022-22501.

Yang, D., Cai, L., Jamieson, S. and Girdhar, Y. 2023. Robot Goes Fishing: Rapid, High-Resolution Biological Hotspot Mapping in Coral Reefs with Vision-Guided Autonomous Underwater Vehicles. Available at: https://arxiv.org/abs/2305.02330.

Yasruddin, M.L., Ismail, M.A.H., Husin, Z. and Tan, W.K. 2022. Feasibility Study of Fish Disease Detection using Computer Vision and Deep Convolutional Neural Network (DCNN) Algorithm. In: *2022 IEEE 18th International Colloquium on Signal Processing & Applications (CSPA)*. pp. 272–276. doi: 10.1109/CSPA55076.2022.9782020.

Yoccoz, N.G., Nichols, J.D. and Boulinier, T. 2001. Monitoring of biological diversity in space and time. *Trends in Ecology & Evolution* 16(8), pp. 446–453. doi: 10.1016/S0169-5347(01)02205-4.

Yosinski, J., Clune, J., Bengio, Y. and Lipson, H. 2014. How transferable are features in deep neural networks? Available at: http://arxiv.org/abs/1411.1792 [Accessed: 17 April 2025].

Zhuang, F. et al. 2020. A Comprehensive Survey on Transfer Learning. Available at: http://arxiv.org/abs/1911.02685 [Accessed: 17 April 2025].

SUPPLEMENTARY MATERIAL

SUPPLEMENTARY MATERIAL, CHAPTER 3

Supplementary Table 1. Search terms used in the R package 'Photosearcher' to investigate the number of records on Flickr of diseases obtained from the World Organisation for Animal Health animal diseases portal.

Disease Name	**Associated Search Terms (comma separated)**
Abalone viral ganglioneuritis	abalone herpesvirus
Acute hepatopancreatic necrosis disease	Vibrio parahaemolyticus
African horse sickness	african horse sickness virus
African swine fever	african swine fever virus
Agent causing chronic wasting disease	CWD prions
Anthrax	Bacillus anthracis
Atrophic rhinitis of swine	Pasteurella multocida
Aujeszkyís disease	Suid herpesvirus 1 (SHV-1)
Avian chlamydiosis	Chlamydia psittaci, Chlamydia avium, Chlamydia gallinacea
Avian infectious bronchitis	Infectuous bronchitis virus
Avian infectious laryngotracheitis	gallid alphaherpesvirus 1
Avian influenza	Bird flu, Orthomyxoviridae Alphainfluenzavirus
Avian mycoplasmosis	Mycoplasma gallisepticum, Mycoplansma synoviae
Avian tuberculosis	Mycobacterium
Babesiosis	Babesia
Baylisascariasis	Baylisascaris procyonis
Bluetongue	Bluetongue virus

Bonamiosis	Bonamia exitiosa, Bonamia ostreae
Border disease	Border disease virus
Lyme disease	Borrelia burgdorferi
Tickborne Relapsing Fever	Borrelia anserina
Bovine anaplasmosis	Anaplasma marginale
Bovine babesiosis	Babesia bovi, Babesia bigemina, Babesia divergens
Bovine genital campylobacteriosis	Campylobacter fetus subsp. venerealis
Bovine spongiform encephalopathy	Bovine spongiform encephalopathy prions
Bovine tuberculosis	Mycobacterium tuberculosis complex
Bovine viral diarrhoea	Bovine viral diarrhoea virus
Brucellosis	Brucella abortus, Brucella melitensis, Brucella suis
Cache Valley virus disease	Cache Valley virus
Akabane disease	Akabane virus
Camelpox	Camelpox virus
Campylobacteriosis	Campylobacter jejuni, Campylobacter coli
Caprine arthritis	Caprine encephalitis, CAE virus
Carp edema virus infection	Carp edema virus
Chytridiomycosis	Batrachochytrium dendrobatidis, Batrachochytrium salamandrivorans
Circovirosis	Canine circovirus, Fox circovirus, Porcine circovirus 1, Porcine circovirus 2, Porcine circovirus 3, Canary circovirus, Goose circovirus, Gull circovirus, Pigeon circovirus, Beak and feather disease, Beak and feather disease virus
Classical swine fever	Classical swine fever virus
Contagious agalactia	Mycoplasma agalactiae, Mycoplasma capricolum subsp. capricolum, Mycoplasma mycoides subsp. capri, Mycoplasma putrefaciens
Contagious bovine pleuropneumonia	Mycoplasma Mycoides
Contagious caprine pleuropneumonia	Mycoplasma capricolum subsp. Capripneumoniae

Contagious equine metritis	Taylorella equigenitalis
Crayfish plague	Aphanomyces astaci
Crimean Congo haemorrhagic fever	Crimean-Congo haemorrhagic fever virus
Cryptosporidiosis	Cryptosporidium
Cysticercosis	Taenia spp.
Decapod iridescent virus 1 infection	Decapod iridescent virus 1
Dermatophilosis	Dermatophilus congolensis
Dourine	Trypanosoma (Trypanozoon) equiperdum
Duck virus enteritis	Anatid alphaherpesvirus-1
Duck virus hepatitis	duck hepatitis A virus, duck astrovirus type 1, duck astrovirus type 2
Ebola Virus Disease	Ebolavirus
Echinococcosis	Echinococcus
Elephant endotheliotropic herpesvirus disease	Elephant Endotheliotropic Herpesvirus, EEHV
Encephalomyocarditis virus infection	Encephalomyocarditis virus
Enzootic abortion of ewes	Chlamydia abortus
Enzootic bovine leukosis	bovine leukaemia virus
Epizootic haematopoietic necrosis disease	epizootic haematopoietic necrosis virus
Epizootic haemorrhagic disease	epizootic haemorrhagic disease virus
Epizootic lymphangitis	Histoplasma capsulatum var. farciminosum
Epizootic ulcerative syndrome	Aphanomyces invadans
Equine encephalomyelitis	Eastern encephalomyelitis virus, Western encephalomyelitis virus
Equine infectious anaemia	Equine Infectious Anaemia virus
Equine influenza	Equine influenza virus (Orthomyxoviridae influenzavirus A subtypes H7N7 and H3N8)
Equine piroplasmosis	Theileria equi, Babesia caballi

Equine viral arteritis	Equine viral arteritis virus
Equine viral rhinopneumonitis	Equid herpesvirus-1 (EHV-1)
European brown hare syndrome	European brown hare syndrome virus
Fasciolosis	Fasciola gigantica, Fascioloides magna
Feline leukaemia	Feline leukaemia virus
Fibropapillomatosis	chelonid herpesvirus 5 (ChHV5)
Filovirosis	filoviruses
Foot and mouth disease	Foot and mouth disease virus
Fowl cholera	Pasteurella multocida
Fowl pox	fowlpox virus
Fowl typhoid	Salmonella Gallinarum biovar Gallinarum
Glanders	Burkholderia mallei
Gyrodactylosis	Gyrodactylus salaris
Haemorrhagic septicaemia	Pasteurella Multocida serotypes B:2 and E:2 (Carter and Heddleston classification system) or 6:B and 6:E (Namioka-Carter classification system)
Hantavirosis	Hantavirus
Heartwater	Ehrlichia ruminantium
Hendra	Hendra virus
Herpesvirosis	Alcelaphine herpesvirus 1, ovine herpesvirus 2
Immunodeficiency virus infection	Feline immunodeficiency virus, FIV, simian immunodeficiency virus, SIV
Covert mortality nodavirus	CMNV
Enterocytozoon hepatopenaei	Enterocytozoon hepatopenaei
Louping ill	Flavivirus, Louping ill virus
Salmon anaemia virus infection	salmon anaemia virus, pathogenic highly polymorphic region (HPR)-deleted, non-pathogenic highly polymorphic region (HPR0)-deleted
Tilapia lake virus infection	Tilapia lake virus

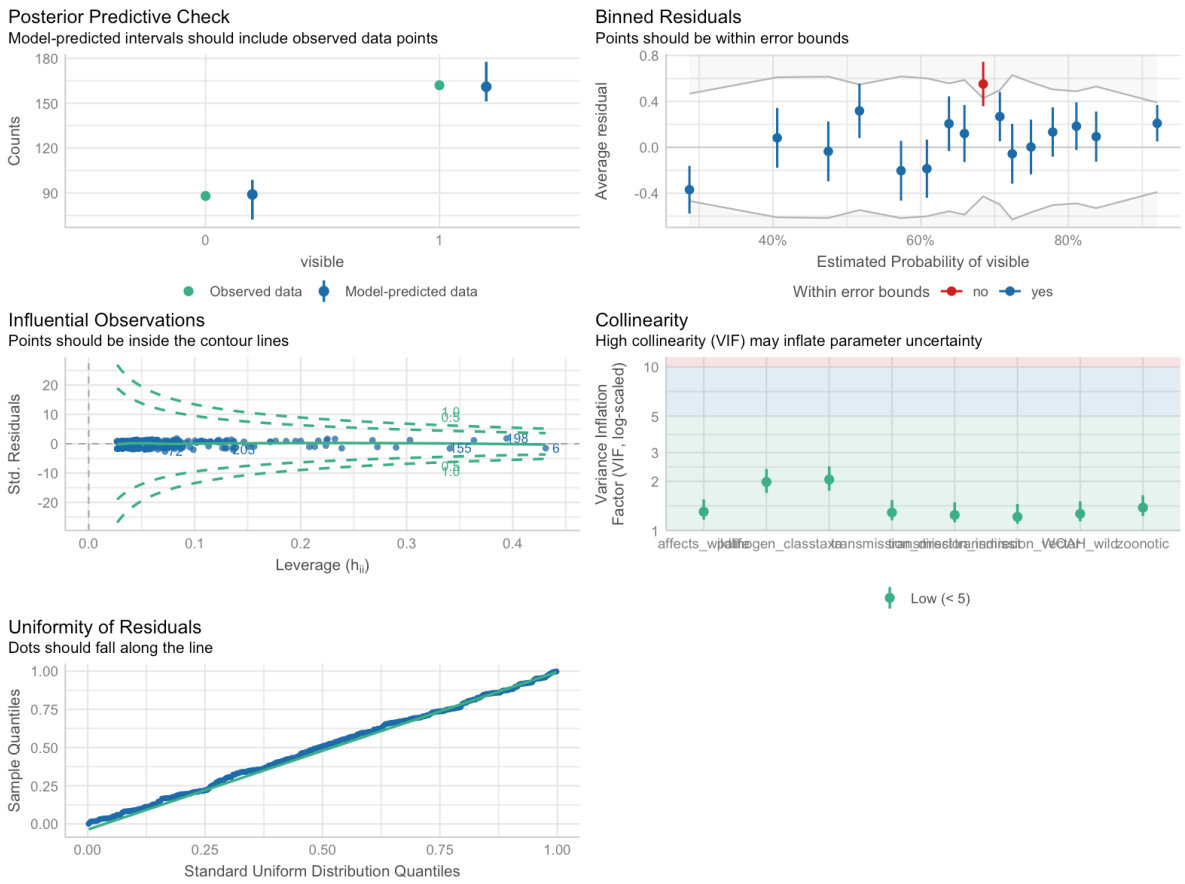
Infectious bovine rhinotracheitis	Infectious pustular vulvovaginitis, bovine herpesvirus 1
Infectious bursal disease	Infectious bursal disease (IBD) virus (Birnaviridae Avibirnavirus) serotype 1
Infectious haematopoietic necrosis	salmonid novirhabdovirus
Hypodermal necrosis	Haematopoietic necrosis, Decapod penstylhamaparvovirus 1
Infectious myonecrosis	infectious myonecrosis virus
Influenza A virus of swine	Influenza A viruses of swine
Japanese encephalitis	Japanese encephalitis virus
Koi herpesvirus disease	cyprinid herpesvirus-3 (CyHV-3)
Leishmaniosis	Leishmania
Leptospirosis	Leptospira, Leptospira interrogans
Listeriosis	Listeria monocytogenes
Low pathogenic avian influenza	Low pathogenic avian influenza viruses
Lumpy Skin Disease	Lumpy skin disease virus
Maedi-visna	maedi-visna virus
Malaria	Plasmodium spp., Plasmodium
Malignant Catarrhal Fever	Alcelaphine gammaherpesvirus-1, Ovine gammaherpesvirus-2
Mange	mange mites
Marekís disease	Marek's disease, Marekís disease virus, Marek's disease virus
Marteiliosis	Marteilia refringens
Middle East respiratory syndrome	MERS, Middle East respiratory syndrome coronavirus (MERS-CoV)
Morbillivirosis	Morbilliviruses
Mpox	Monkeypox virus
Myxomatosis	Myxoma virus
Nairobi sheep disease	Nairobi sheep disease virus

Necrotising hepatopancreatitis	Candidatus Hepatobacter penaei
New world screwworm	Cochliomyia hominivorax
Newcastle disease	Newcastle disease virus
Nipah	Nipah virus
Nosemosis	Nosema apis (Zander), Nosema ceranae, Nosema
Old world screwworm	Chrysomya bezziana Villeneuve
Ovine epididymitis	Brucella ovis
Ovine pulmonary adenocarcinoma	jaagsiekte sheep retrovirus
Papillomatosis in crocodiles	crocodile poxvirus
Paramyxovirosis	Avian paramyxoviruses
Paratuberculosis	Mycobacterium avium subsp. paratuberculosis
Parvovirosis	Parvoviruses, Hepatopancreatic parvovirus, infectious hypodermal and hematopoietic necrosis virus, Densoviruses
Pasteurellosis	Pasteurella spp
Perkinsosis	Perkinsus marinus, Perkinsus olseni
Peste des petits ruminants	small ruminant Morbillivirus
Porcine cysticercosis	Taenia solium
Porcine epidemic diarrhoea	Porcine epidemic diarrhoea virus
Porcine reproductive and respiratory syndrome	Porcine reproductive and respiratory syndrome virus
Pox viruses infection	Pox, Poxviruses
Psoroptic mange	Psoroptes spp.
Pullorum disease	Salmonella enterica subspecies enterica serovar Gallinarum biovar Pullorum
Q fever	Coxiella burnetii
Rabbit haemorrhagic disease	RHDB, calicivirus
Rabies	Rabies virus, Lyssavirus
Ranavirosis	Ranavirus, Ranaviruses

Red sea bream iridoviral disease	RSIVD, Red sea bream iridovirus
Rift Valley fever	Rift valley fever virus
Rinderpest	Rinderpest virus
Salmonellosis	Salmonella spp., Salmonella abortusovis, Salmonella enterica
Salmonid alphavirus infection	Salmonid alphavirus
Scabies	Sarcoptes scabiei mite
Schmallenberg disease	Schmallenberg virus
Scrapie	PrPc -> PrPsc
Sheep pox	Goat pox, Sheeppox virus, Goatpox virus
Snake fungal disease	Ophidiomyces ophiodiicola
Spring viraemia of carp	Carp sprivivirus
Surra	Trypanosoma evansi
Swine vesicular disease	Swine vesicular disease virus
Taura syndrome	Taura syndrome virus
Teschovirus encephalomyelitis	porcine teschovirus serotype 1 (PTV-1)
Theileriosis	Theileria annulata, Theileria parva, Theileria orientalis, Theileria spp.
Tick borne encephalitis	TBE, Tickborne encephalitis virus
Toxoplasmosis	Toxoplasma gondii
Transmissible gastroenteritis	Transmissible gastroenteritis virus
Trichinellosis	Trichinella spp., Trichinella britovi, Trichinella murrelli, Trichinella nativa, Trichinella nelsoni, Trichinella papuae, Trichinella patagoniensis, Trichinella pseudospiralis, Trichinella spiralis
Trichomonosis	Tritrichomonas foetus, Trichomonas spp.
Trypanosomiase	Order Kinetoplastida; family Trypanosomatidae; Genus Trypanosoma; Subgenus Nannomonas (T. congolense), Subgenus Duttonella (T. vivax), and Subgenus Trypanozoon (T. brucei ssp.).

Tularemia	Francisella tularensis subsp. tularensis, Francisella tularensis subsp. holarctica
Turkey rhinotracheitis	Avian metapneumovirus (aMPV)
Venezuelan equine encephalitis	Venezuelan equine encephalomyelitis (VEE) viruses
Verocytotoxigenic Escherichia coli	Verocytotoxigenic Escherichia coli
Vesicular Stomatitis	vesiculoviruses
Viral haemorrhagic septicaemia	viral haemorrhagic septicaemia virus
West Nile fever	West Nile virus
White spot disease	white spot syndrome virus
White tail disease	Macrobrachium rosenbergii nodavirus (MrNV)
White-nose syndrome	Pseudogymnoascus destructans
Withering abalone syndrome	Xenohaliotis californiensis
Yellow fever	Yellow fever virus
Yellow head disease	yellow head virus genotype 1 (YHV1)
Yersiniosis enterocolitica	Yersinia enterocolitica
Yersiniosis pestis	Plague, Yersinia pestis
Yersiniosis pseudotuberculosis	Yersinia pseudotuberculosis
Foamy virus	Simian foamy viruses
Hepatitis A	Simian hepatitis A virus
Hepatitis B	Hepatitis B virus
Herpes	Saimiriine Herpesvirus 1
Herpesvirus Cercopithecus infection	Cercopithecine herpesvirus 2
Herpesvirus saimiri infection	Saimiriine herpesvirus 2
Macacine herpesvirus infection	Macacine herpesvirus 1 (McHV1)
Marburg virus disease	Marburg virus
Papiine herpesvirus 2 infection	Papiine herpesvirus 2
Simian haemorrhagic fever	Simian haemorrhagic fever virus
Simian retrovirus infection	Simian betaretroviruses, Simian retroviruses

Simian T-cell lymphotropic virus-1 infection	Simian T-cell lymphotrophic virus-1
SV40 infection	Simian Virus 40 (SV40)
Zika	Zika virus
Burkholderia pseudomallei	Burkholderia pseudomallei
Shigella flexneri infection	Shigella flexneri
Entamoeba histolytica infection	Entamoeba histolytica
Giardia infection	Giardia spp.
Pneumonyssus simicola infestation	Pneumonyssus simicola
Prosthenorchis elegans infection	Prosthenorchis elegans
Strongyloides stercoralis infection	Strongyloides stercoralis
Trichuriasis	Trichuris
Trichophyton infection	Trichophyton sp.
Acarapisosis	Acarapis woodi
American foulbrood	Paenibacillus larvae
European foulbrood	Melissococcus plutonius
Small hive beetle infestation	Aethina tumida
Tropilaelaps	Tropilaelaps spp.
Varroosis	Varroa, Varroa destructor

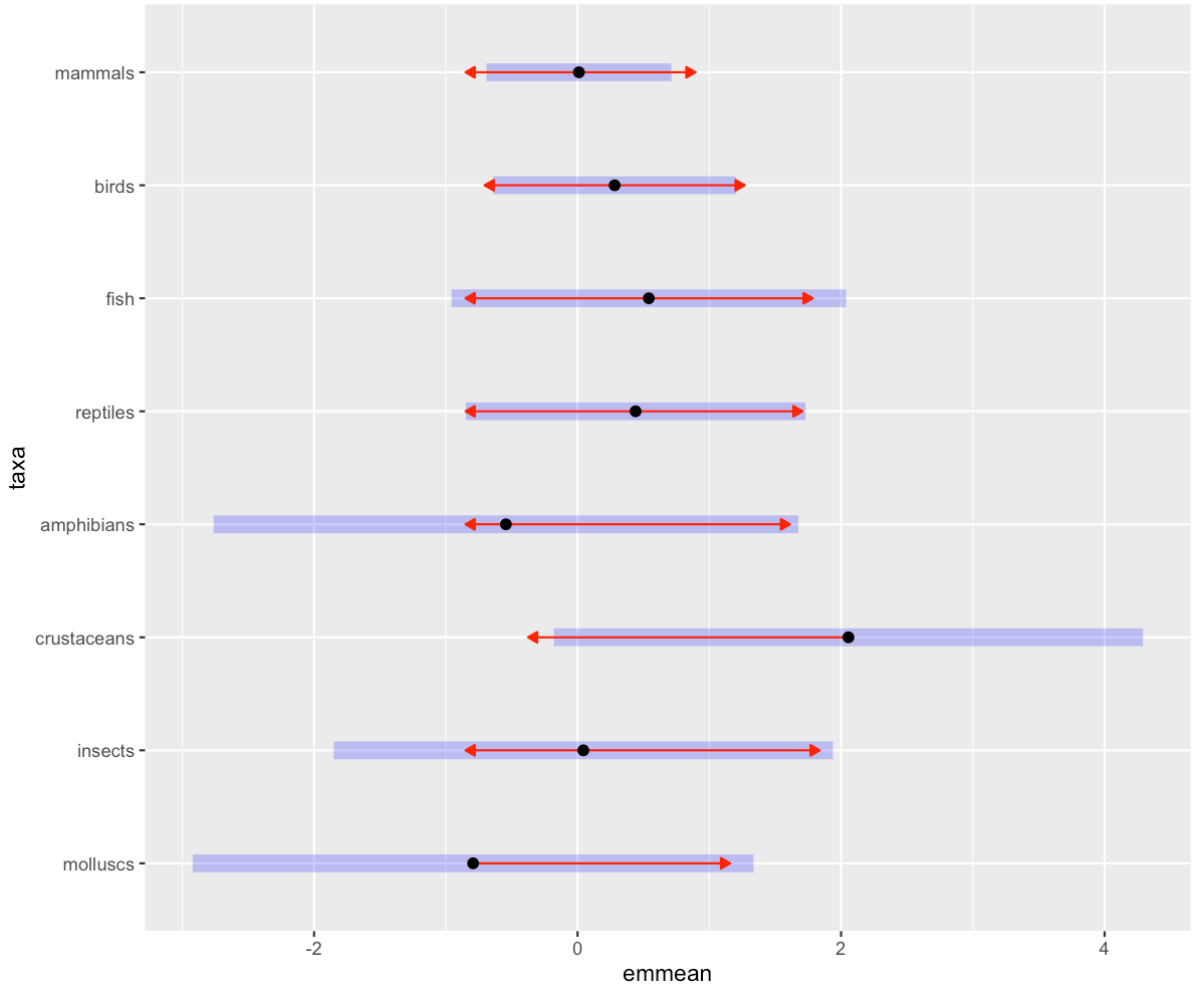


Supplementary Figure 1. Model checks returned by the R package 'performance' (Lüdecke et al. 2021) for a binomial glm used to investigate the association between taxa, zoonotic status, wild status, and pathogen class, with the occurrence of visible disease signs obtained from the World Organisation for Animal Health animal diseases portal.

Supplementary Table 2. Coefficients from the binomial Generalised Linear Model used to predict the presence of visible disease signs. The reference levels were 'mammals' for Host Taxa and 'virus' for Pathogen Class. The asterisk (*) denotes statistical significance ($p < 0.05$).

Variable	Level	Estimate	Std. Error	z value	Pr(>\z\)
(Intercept)		-0.61	0.51	-1.2	0.23
taxa	Molluscs	-0.8	1.03	-0.78	0.43357

	Insects	0.03	0.96	0.03	0.97209
	Crustaceans	2.05	1.1	1.86	0.0634
	Amphibians	-0.55	1.11	-0.5	0.61609
	Reptiles	0.43	0.63	0.69	0.49293
	Fish	0.53	0.67	0.79	0.43023
	Birds	0.27	0.39	0.69	0.49117
zoonotic1	Yes	-0.33	0.33	-1	0.317
wild_alt_source1	Yes	0.66	0.41	1.6	0.11
WOAH_wild1	Yes	-0.23	0.35	-0.67	0.504
	Prion	-0.98	1.32	-0.74	0.45807
	Bacterium	0.06	0.38	0.16	0.86992
	Protist	0.03	0.46	0.07	0.94078
	Ectoparasite	0.7	0.87	0.81	0.41876
	Fungi	0.04	0.94	0.05	0.96344
pathogen_class	Helminth	-0.9	0.71	-1.27	0.20264
transmission_direct1	Yes	0.94	0.33	2.87	0.004*
transmission_indirect1	Yes	0.15	0.33	0.46	0.648
transmission_vector1	Yes	0.36	0.35	1.01	0.311



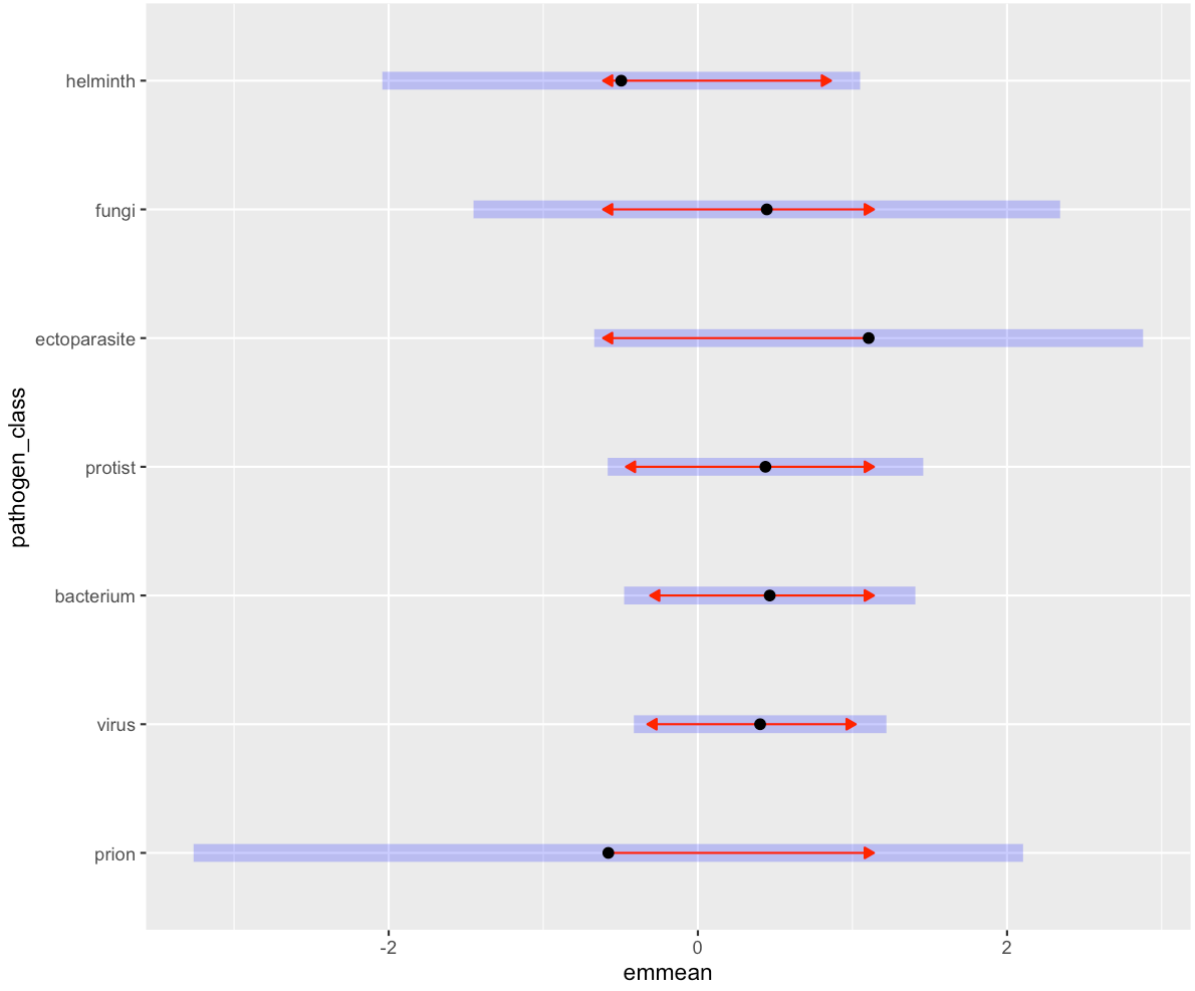
Supplementary Figure 2. Post-hoc pairwise comparisons returned by the R package 'emmeans' (Lenth 2024) of the tendency for different taxonomic groups to show visible disease signs for diseases from the World Organisation for Animal Health animal diseases portal.

Supplementary Table 3. Post-hoc pairwise comparisons of the tendency for different host taxa to show visible disease signs for diseases in the World Organisation for Animal Health

animal diseases portal, from the R package emmeans (Lenth 2024). The asterisk (*) denotes statistical significance ($p < 0.05$).

contrast	estimate	SE	df	z.ratio	p.value
mammals - molluscs	0.8	1.03	Inf	0.78	0.99
mammals - insects	-0.03	0.96	Inf	-0.03	1
mammals - crustaceans	-2.05	1.1	Inf	-1.86	0.58
mammals - amphibians	0.55	1.11	Inf	0.5	1
mammals - reptiles	-0.43	0.63	Inf	-0.69	1
mammals - fish	-0.53	0.67	Inf	-0.79	0.99
mammals - birds	-0.27	0.39	Inf	-0.69	1
molluscs - insects	-0.84	1.33	Inf	-0.63	1
molluscs - crustaceans	-2.85	1.44	Inf	-1.98	0.49
molluscs - amphibians	-0.25	1.42	Inf	-0.18	1
molluscs - reptiles	-1.23	1.15	Inf	-1.08	0.96
molluscs - fish	-1.33	1.13	Inf	-1.18	0.94
molluscs - birds	-1.07	1.06	Inf	-1.01	0.97
insects - crustaceans	-2.01	1.4	Inf	-1.44	0.84
insects - amphibians	0.59	1.39	Inf	0.42	1
insects - reptiles	-0.4	1.1	Inf	-0.36	1
insects - fish	-0.5	1.11	Inf	-0.45	1
insects - birds	-0.24	1	Inf	-0.24	1
crustaceans - amphibians	2.6	1.51	Inf	1.72	0.68
crustaceans - reptiles	1.62	1.23	Inf	1.31	0.9
crustaceans - fish	1.52	1.23	Inf	1.24	0.92
crustaceans - birds	1.77	1.14	Inf	1.56	0.77

amphibians - reptiles	-0.98	1.21	Inf	-0.81	0.99
amphibians - fish	-1.08	1.23	Inf	-0.88	0.99
amphibians - birds	-0.83	1.14	Inf	-0.72	1
reptiles - fish	-0.1	0.87	Inf	-0.11	1
reptiles - birds	0.16	0.68	Inf	0.23	1
fish - birds	0.26	0.74	Inf	0.35	1



Supplementary Figure 3. Post-hoc pairwise comparisons returned by the R package 'emmeans' (Lenth 2024) of the tendency for different pathogen classes to show visible

disease signs for diseases from the World Organisation for Animal Health animal diseases portal

Supplementary Table 4. Post-hoc pairwise comparisons of the tendency for different pathogens to show visible disease signs for diseases in the World Organisation for Animal Health animal diseases portal, from the R package emmeans (Lenth 2024).

contrast	**estimate**	**SE**	**df**	**z.ratio**	**p.value**
virus - prion	0.98	1.32	Inf	0.74	0.99
virus - bacterium	-0.06	0.38	Inf	-0.16	1
virus - protist	-0.03	0.46	Inf	-0.07	1
virus - ectoparasite	-0.7	0.87	Inf	-0.81	0.98
virus - fungi	-0.04	0.94	Inf	-0.05	1
virus - helminth	0.9	0.71	Inf	1.27	0.86
prion - bacterium	-1.04	1.34	Inf	-0.78	0.99
prion - protist	-1.02	1.38	Inf	-0.74	0.99
prion - ectoparasite	-1.68	1.58	Inf	-1.07	0.94
prion - fungi	-1.03	1.6	Inf	-0.64	1
prion - helminth	-0.08	1.47	Inf	-0.06	1
bacterium - protist	0.03	0.51	Inf	0.05	1
bacterium - ectoparasite	-0.64	0.89	Inf	-0.72	0.99
bacterium - fungi	0.02	0.97	Inf	0.02	1
bacterium - helminth	0.96	0.72	Inf	1.33	0.84
protist - ectoparasite	-0.67	0.96	Inf	-0.7	0.99
protist - fungi	-0.01	0.99	Inf	-0.01	1
protist - helminth	0.93	0.77	Inf	1.22	0.89
ectoparasite - fungi	0.66	1.22	Inf	0.54	1

ectoparasite - helminth	1.6	1.09	Inf	1.47	0.76
fungi - helminth	0.94	1.13	Inf	0.83	0.98

Supplementary Table 5. The number of records returned by the R package 'Photosearcher' for diseases obtained from the World Organisation for Animal Health animal diseases portal.

Disease name	**Number of images**
Malaria	5018
Anthrax	3789
Pox viruses infection	3147
Yellow fever	2046
Border disease	1281
Avian influenza	1216
Hendra	977
Herpes	562
Bluetongue	533
Newcastle disease	487
White spot disease	393
Lyme disease	351
Ebola Virus Disease	339
Psoroptic mange	290

Chytridiomycosis	269
White tail disease	251
Yellow head disease	248
Scabies	219
Marburg virus disease	210
Foot and mouth disease	186
Circovirosis	166
West Nile fever	157
Varroosis	142
Zika	130
White-nose syndrome	116
Myxomatosis	106
Surra	97
Giardia infection	93
Rinderpest	63
Trichomonosis	54
Rift Valley fever	49
Influenza A virus of swine	42
African swine fever	36
Avian tuberculosis	29
Snake fungal disease	29

Leptospirosis	29
African horse sickness	25
Leishmaniosis	21
Toxoplasmosis	19
New world screwworm	14
Cryptosporidiosis	11
Sheep pox	10
Theileriosis	10
Trichuriasis	10
Salmonellosis	9
Scrapie	9
Mpox	9
Contagious bovine pleuropneumonia	9
Lumpy Skin Disease	8
Pasteurellosis	8
Fowl cholera	8
Atrophic rhinitis of swine	8
Swine vesicular disease	7
Avian mycoplasmosis	7
Vesicular Stomatitis	7

Rabbit haemorrhagic disease	6
Strongyloides stercoralis infection	6
Marek's disease	6
Equine infectious anaemia	5
Fowl pox	5
Gyrodactylosis	5
Fibropapillomatosis	4
Ranavirosis	4
Crayfish plague	4
Elephant endotheliotropic herpesvirus disease	2
Covert mortality nodavirus	2
Shigella flexneri infection	2
Acarapisosis	2
Duck virus enteritis	1
Haemorrhagic septicaemia	1
Paratuberculosis	1
Parvovirosis	1
Peste des petits ruminants	1
Trichophyton infection	1

SUPPLEMENTARY MATERIAL, CHAPTER 4

OBSERVER BIAS

Methods

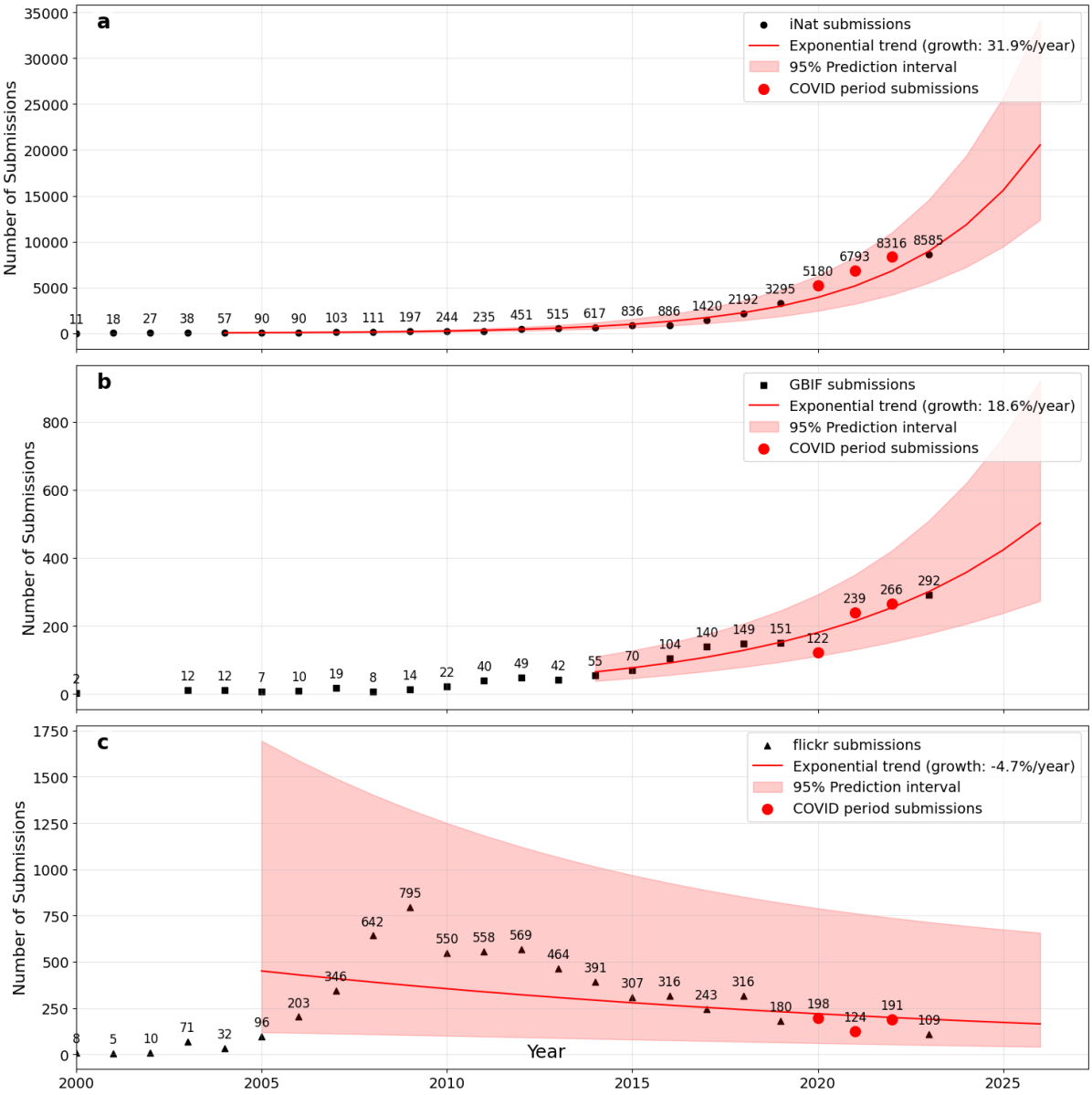
Recent studies show that citizen science activity changed markedly during the COVID-19 pandemic, with many platforms reporting increased participation (Sánchez-Clavijo et al. 2021; Dimson and Gillespie 2023; Qiao et al. 2023). To model long-term growth and assess whether similar patterns were evident in our dataset, we conducted an exponential regression analysis. Wikimedia Commons was excluded from this analysis due to insufficient data. For the remaining platforms, the analysis began in the first year with ≥50 submissions to ensure sufficient data for robust regression while still capturing early growth trends.

To model exponential growth, we log-transformed the annual submission counts. This linearises the growth trend and stabilises variance, a common approach for count data where variance often scales with the mean (Wooldridge 2020). Key assumptions for linear regression, such as normality and homoscedasticity of residuals, were checked and met. We then fitted a simple linear model ($\ln(y) = c + bx$) to the transformed data using ordinary least squares, where y is the submission count and x is the year (normalised to start from zero). The 95% prediction intervals were calculated for this model and subsequently back-transformed to the original scale by taking their exponent, producing the final exponential trend line and prediction bands. The annual growth rate for each platform was calculated from the model's slope parameter b as ($e^b - 1$) * 100% (Wooldridge 2020).

To assess observer bias, we used the COVID-19 pandemic as a case study, testing whether submission counts during the period of major global mobility restrictions fell within the 95% prediction interval of the pre-pandemic model. This period was defined as 11 March 2020 (the WHO pandemic declaration) to 30 June 2022 (when major travel and social restrictions were broadly eased in North America and Europe).

Results

Exponential regression analysis of long-term trends showed annual growth rates of +31.7% for iNaturalist (Supplementary Figure 4a) and +17.7% for GBIF (Supplementary Figure 4b), and -3.7% for Flickr (Supplementary Figure 4c). During the 2020-2022 COVID-19 period, the number of submissions for these platforms fell within their models' 95% prediction intervals, though GBIF submissions in 2020 were 33% lower than expected. Salmonid image submissions rose from 7,506 in the three years preceding the pandemic (2017-2019) to 30,420 during it. iNaturalist supplied 96.5% of these pandemic-era submissions.



Supplementary Figure 4. Exponential trends in salmonid image submissions across digital platforms (2000-2023). (a) iNaturalist submissions. (b) GBIF submissions. (c) Flickr

submissions. Solid lines represent fitted exponential models (models have been fitted from the first year with ≥50 submissions following the platform's launch), shaded areas indicate 95% prediction intervals, blue/purple/green points show yearly submission counts and red points highlight submissions from the period where substantial COVID-19 restrictions were in effect globally (2020-2022).

Discussion

Our analysis of observer bias showed that the increase in submissions during the COVID-19 pandemic, mostly from iNaturalist, matches trends seen in other citizen science projects (Dimson and Gillespie 2023; Qiao et al. 2023). This period shows that citizen science can be a resilient and scalable way to collect data, but also suggests that major events can create temporal biases that must be considered in long-term analyses.

Supplementary Table 6. Taxonomic Classification and keywords of Salmonidae Family used to search photo-sharing sites.

Scientific name (taxonomic rank)	**Common name**
Salmonidae (family)	
Coregoninae (subfamily)	whitefish
Coregonus (genus)	
Coregonus acrinasus (species)	
Coregonus albellus (species)	
Coregonus aff. albellus Brienzer (species)	
Coregonus aff. albellus Thuner (species)	
Coregonus albula (species)	vendace
Coregonus alpenae (species)	longjaw cisco
Coregonus alpinus (species)	

Scientific name (taxonomic rank)	Common name
Coregonus aff. alpinus Brienzer (species)	
Coregonus aff. alpinus Thuner (species)	
Coregonus anaulorum (species)	
Coregonus arenicolus (species)	
Coregonus artedi (species)	cisco
Coregonus atterensis (species)	
Coregonus austriacus (species)	
Coregonus autumnalis (species)	arctic cisco
Coregonus baerii (species)	
Coregonus baicalensis (species)	baikal whitefish
Coregonus baunti (species)	
Coregonus bavaricus (species)	
Coregonus bezola (species)	
Coregonus brienzii (species)	
Coregonus candidus (species)	
Coregonus chadary (species)	khadary whitefish
Coregonus clupeaformis (species)	lake whitefish
Coregonus clupeoides (species)	powan
Coregonus confusus (species)	
Coregonus danneri (species)	
Coregonus duplex (species)	
Coregonus fatioi (species)	
Coregonus cf. *duplex* AGH-2010 (species)	
Coregonus fatioi (species)	
Coregonus fera (species)	
Coregonus fluviatilis (species)	

Scientific name (taxonomic rank)	Common name
Coregonus fontanae (species)	
Coregonus gutturosus (species)	lake constance whitefish
Coregonus heglingus (species)	
Coregonus aff. heglingus Walen (species)	
Coregonus aff. heglingus Zurich (species)	
Coregonus hiemalis (species)	
Coregonus hoferi (species)	
Coregonus holsata (species)	
Coregonus hoyi (species)	bloater
Coregonus huntsmani (species)	atlantic whitefish
Coregonus intermundia (species)	
Coregonus johannae (species)	deepwater cisco
Coregonus kiletz (species)	
Coregonus kiyi (species)	kiyi
Coregonus ladogae (species)	
Coregonus laurettae (species)	bering cisco
Coregonus lavaretus (species)	european whitefish; common whitefish
Coregonus litoralis (species)	
Coregonus lucinensis (species)	
Coregonus lutokka (species)	
Coregonus macrophthalmus (species)	
Coregonus maraena (species)	maraena whitefish
Coregonus maraenoides (species)	
Coregonus maxillaris (species)	

Scientific name (taxonomic rank)	Common name
Coregonus megalops (species)	lacustrine fluvial whitefish
Coregonus migratorius (species)	arctic cisco
Coregonus muelleri (species)	
Coregonus muksun (species)	muksun
Coregonus nasus (species)	broad whitefish
Coregonus nelsonii (species)	alaska whitefish
Coregonus nigripinnis (species)	blackfin cisco
Coregonus nilssoni (species)	
Coregonus nipigon (species)	
Coregonus nobilis (species)	
Coregonus oxyrinchus (species)	houting
Coregonus palaea (species)	
Coregonus aff. palaea Biel (species)	
Coregonus aff. palaea Neuchatel (species)	
Coregonus pallasii (species)	
Coregonus peled (species)	peled
Coregonus pennantii (species)	gwyniad
Coregonus pidschian (species)	humpback whitefish
Coregonus pollan (species)	irish pollan
Coregonus pravdinellus (species)	
Coregonus profundus (species)	
Coregonus reighardi (species)	shortnose cisco
Coregonus renke (species)	
Coregonus restrictus (species)	
Coregonus sardinella (species)	sardine cisco; least cisco

Scientific name (taxonomic rank)	Common name
Coregonus steinmanni (species)	
Coregonus stigmaticus (species)	schelly
Coregonus subautumnalis (species)	
Coregonus suidteri (species)	
Coregonus cf. *suidteri* AGH-2010 (species)	
Coregonus cf. *suidteri* Ex74D1 (species)	
Coregonus cf. *suidteri* Ex74D2 (species)	
Coregonus cf. *suidteri* Ex74D3 (species)	
Coregonus cf. *suidteri* Ex74D4 (species)	
Coregonus cf. *suidteri* Ex74D5 (species)	
Coregonus cf. *suidteri* Ex74D6 (species)	
Coregonus cf. *suidteri* Ex74E1 (species)	
Coregonus cf. *suidteri* Ex74E2 (species)	
Coregonus cf. *suidteri* Ex74E3 (species)	
Coregonus cf. *suidteri* Ex74E4 (species)	
Coregonus cf. *suidteri* Ex74E5 (species)	
Coregonus cf. *suidteri* Ex74E6 (species)	
Coregonus cf. *suidteri* LJDQ-2022 (species)	
Coregonus suspensus (species)	
Coregonus trybomi (species)	
Coregonus tugun (species)	tugun
Coregonus ussuriensis (species)	amur whitefish
Coregonus vandesius (species)	
Coregonus vessicus (species)	
Coregonus wartmanni (species)	blaufelchen
Coregonus widegreni (species)	valaam whitefish

Scientific name (taxonomic rank)	**Common name**
Coregonus zenithicus (species)	shortjaw cisco
Coregonus zuerichensis (species)	
Coregonus cf. *zuerichensis* AGH-2010 (species)	
Coregonus zugensis (species)	
Prosopium (genus)	freshwater whitefish
Prosopium abyssicola (species)	bear lake whitefish
Prosopium coulterii (species)	pygmy whitefish
Prosopium cylindraceum (species)	round whitefish
Prosopium gemmifer (species)	bonneville cisco
Prosopium spilonotus (species)	bonneville whitefish
Prosopium williamsoni (species)	mountain whitefish
Stenodus (genus)	
Stenodus leucichthys (species)	sheefish; inconnu
Stenodus nelma (species)	
Salmoninae (subfamily)	trout; salmon; char
Brachymystax (genus)	
Brachymystax lenok (species)	lenok; lenok trout
Brachymystax savinovi (species)	
Brachymystax tsinlingensis (species)	
Brachymystax tumensis (species)	
Hucho (genus)	
Hucho bleekeri (species)	sichuan taimen
Hucho hucho (species)	huchen
Hucho ishikawae (species)	
Hucho perryi (species)	
Hucho taimen (species)	taimen

Scientific name (taxonomic rank)	**Common name**
Oncorhynchus (genus)	
Oncorhynchus aguabonita (species)	golden trout
Oncorhynchus apache (species)	apache trout
Oncorhynchus chrysogaster (species)	mexican golden trout
Oncorhynchus clarkii (species)	cutthroat trout
Oncorhynchus formosanus (species)	
Oncorhynchus gilae (species)	gila trout
Oncorhynchus gorbuscha (species)	pink salmon
Oncorhynchus iwame (species)	iwame trout
Oncorhynchus kawamurae (species)	
Oncorhynchus keta (species)	chum salmon
Oncorhynchus kisutch (species)	coho salmon
Oncorhynchus masou (species)	cherry salmon
Oncorhynchus mykiss (species)	rainbow trout
Oncorhynchus nerka (species)	sockeye salmon
Oncorhynchus rhodurus (species)	japanese amago
Oncorhynchus tshawytscha (species)	chinook salmon
Oncorhynchus tshawytscha x *Oncorhynchus kisutch* (hybrid)	
Parahucho (genus)	
Parahucho perryi (species)	japanese huchen
Salmo (genus)	
Salmo abanticus (species)	abant trout
Salmo akairos (species)	
Salmo aphelios (species)	
Salmo balcanicus (species)	
Salmo baliki (species)	

Scientific name (taxonomic rank)	Common name
Salmo carpio (species)	
Salmo caspius (species)	caspian trout
Salmo cenerinus (species)	
Salmo cenerinus 637 HS-2020 (species)	
Salmo cenerinus 640 HS-2020 (species)	
Salmo cenerinus BOLD:AAB3872 (species)	
Salmo cettii (species)	mediterranean Trout
Salmo chilo (species)	
Salmo ciscaucasicus (species)	caspian salmon
Salmo coruhensis (species)	coruh trout
Salmo dentex (species)	
Salmo euphrataeus (species)	euphrates trout
Salmo ezenami (species)	
Salmo fahrettini (species)	
Salmo farioides (species)	balkan brook trout
Salmo ferox (species)	
Salmo fibreni (species)	
Salmo ghigii (species)	
Salmo ischchan (species)	sevan trout
Salmo kottelati (species)	antalya trout
Salmo labecula (species)	seyhan trout
Salmo labrax (species)	black sea salmon
Salmo letnica (species)	ohrid trout; lake ohrid brown trout
Salmo lourosensis (species)	
Salmo lumi (species)	
Salmo macedonicus (species)	macedonian trout

Scientific name (taxonomic rank)	Common name
Salmo macrostigma (species)	
Salmo marmoratus (species)	marble Trout
Salmo marmoratus x *Salmo trutta* (hybrid)	
Salmo montenigrinus (species)	
Salmo munzuricus (species)	
Salmo multipunctatus (species)	draa trout
Salmo nigripinnis (species)	
Salmo obtusirostris (species)	adriatic trout
Salmo ohridanus (species)	
Salmo okumusi (species)	okumus trout
Salmo opimus (species)	opimus trout
Salmo oxianus (species)	
Salmo pallaryi (species)	lake sidi ali trout
Salmo pelagonicus (species)	
Salmo pellegrini (species)	
Salmo peristericus (species)	prespa trout
Salmo platycephalus (species)	flathead trout
Salmo rhodanensis (species)	rhone trout
Salmo rizeensis (species)	rize trout
Salmo salar (species)	atlantic salmon
Salmo schiefermuelleri (species)	
Salmo stomachicus (species)	
Salmo taleri (species)	
Salmo tigridis (species)	tigros trout
Salmo trutta (species)	river trout; brown trout; sea trout
Salmo viridis (species)	

Scientific name (taxonomic rank)	Common name
Salmo visovacensis (species)	
Salmo zrmanjaensis (species)	
Salmo abanticus complex sp. H1 (species)	
Salmo abanticus complex sp. H2 (species)	
Salmo caspius complex sp. H1 (species)	
Salmo caspius complex sp. H2 (species)	
Salmo trutta complex sp. GP-2022 (species)	
Salmo trutta complex sp. Lt-2021 (species)	
Oncorhynchus mykiss x *Salmo salar* (hybrid)	
Salmo trutta x *Salvelinus* (hybrid)	
Salvelinus (genus)	
Salvelinus agassizii (species)	silver trout
Salvelinus albus (species)	white char
Salvelinus alpinus (species)	atlantic char
Salvelinus anaktuvukensis (species)	angayukaksurak char
Salvelinus andriashevi (species)	chukot char
Salvelinus boganidae (species)	boganida char
Salvelinus colii (species)	
Salvelinus confluentus (species)	bull trout
Salvelinus curilus (species)	
Salvelinus czerskii (species)	
Salvelinus drjagini (species)	dryanin's char
Salvelinus elgyticus (species)	small-mouth char
Salvelinus evasus (species)	
Salvelinus faroensis (species)	
Salvelinus fimbriatus (species)	

Scientific name (taxonomic rank)	Common name
Salvelinus fontinalis (species)	brook trout
Salvelinus fontinalis x *Salvelinus malma* (hybrid)	
Salvelinus gracillimus (species)	
Salvelinus grayi (species)	
Salvelinus gritzenkoi (species)	
Salvelinus inframundus (species)	
Salvelinus jacuticus (species)	yakutian char
Salvelinus japonicus (species)	
Salvelinus killinensis (species)	
Salvelinus krogiusae (species)	
Salvelinus kronocius (species)	
Salvelinus kuznetzovi (species)	
Salvelinus lepechini (species)	
Salvelinus leucomaenis (species)	whitespotted char
Salvelinus levanidovi (species)	
Salvelinus lonsdalii (species)	
Salvelinus mallochi (species)	
Salvelinus malma (species)	dolly varden
Salvelinus malma x *Salvelinus leucomaenis* (hybrid)	
Salvelinus maxillaris (species)	
Salvelinus murta (species)	
Salvelinus namaycush (species)	lake trout
Salvelinus namaycush x *Salvelinus fontinalis* (hybrid)	
Salvelinus neiva (species)	nieva
Salvelinus neocomensis (species)	
Salvelinus obtusus (species)	

Scientific name (taxonomic rank)	Common name
Salvelinus perisii (species)	
Salvelinus profundus (species)	
Salvelinus salvelinoinsularis (species)	
Salvelinus schmidti (species)	
Salvelinus struanensis (species)	
Salvelinus taimyricus (species)	
Salvelinus taranetzi (species)	
Salvelinus thingvallensis (species)	
Salvelinus tolmachoffi (species)	
Salvelinus umbla (species)	
Salvelinus vasiljevae (species)	
Salvelinus willoughbii (species)	
Salvelinus youngeri (species)	
Salvethymus (genus)	
Salvethymus svetovidovi (species)	long-finned charr
Thymallinae (subfamily)	grayling
Thymallus (genus)	
Thymallus aeliani (species)	
Thymallus arcticus (species)	arctic grayling
Thymallus baicalensis (species)	baikal black grayling
Thymallus bailolenensis (species)	
Thymallus brevicephalus (species)	
Thymallus brevipinnis (species)	
Thymallus brevirostris (species)	mongolian grayling
Thymallus burejensis (species)	
Thymallus flavomaculatus (species)	yellow spotted grayling

Scientific name (taxonomic rank)	Common name
Thymallus grubii (species)	amur grayling
Thymallus ligericus (species)	
Thymallus mertensii (species)	
Thymallus nigrescens (species)	kosogol grayling
Thymallus nikolskyi (species)	
Thymallus pallasii (species)	east siberian grayling
Thymallus svetovidovi (species)	
Thymallus thymallus (species)	
Thymallus tugarinae (species)	lower amur grayling
Thymallus yaluensis (species)	

Supplementary Table 7. Metadata field standardisation across online image sources, organised by category. The 'Standard field' describes the name of the field in the consolidated metadata.

* the value of the field was based on search terms (Supplementary Table 6) used to download each image for this source

** the value of the field was based on the label for each image following manual verification for this source

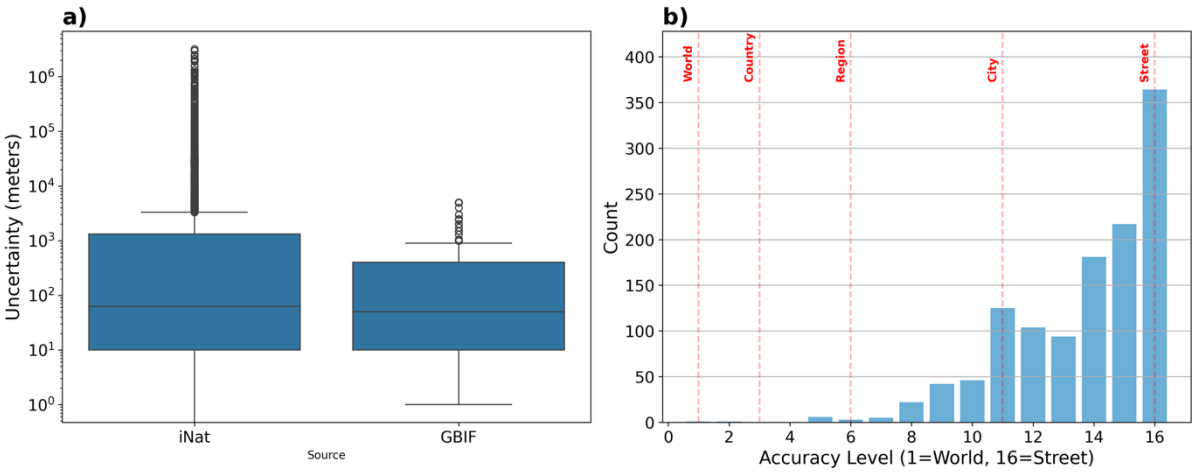
*** the value of the field was based on the unique SHA-256 hash value for each image for this source

**** the value of the field was derived from the Exchangeable Image File Format (EXIF) information for each image for this source

Category	Standard field	iNaturalist	GBIF	Flickr	Wikimedia

Taxonomy	scientific_name	name	verbatimScientificName	based on search term*	based on search term*
	taxonomic_level	rank	taxonRank	based on search term*	based on search term*
	quality	quality_grade	identificationVerificationStatus	label**	label**
Identifier	id_at_source	photo_uuid	identifier	Based on file hash***	image
Temporal	date	observed_on	eventDate	datetaken	Based on EXIF data****
Spatial	latitude	latitude	decimalLatitude	latitude	
	longitude	longitude	decimalLongitude	longitude	
	Location_accuracy	positional_accuracy	coordinateUncertaintyInMeters	accuracy	
Usage rights	license	license	license	license	license
Additional fields, not standardised		observer_id, height, width, ancestry, rank_level, observation_uui, observer_id, position, extension, photo_id	basisOfRecord, institutionCode, collectionCode, datasetName, ownerInstitutionCode, informationWithheld, dataGeneralizations, dynamicProperties, occurrenceID, catalogNumber, recordNumber, recordedBy, recordedByID, kingdom, phylum, class, order, family,	owner_username, owner_path, height, width, url, title, description, tags, urls, usage_canblog, usage_canprint, usage_canshare, has_people, date_uploaded_raw, date_posted_raw, date_taken_raw, date_updated_raw,	page.title, image.url, desc.url, image.desc, user, artist, copyrighted, usage, attribution.req, likely.genus.only, distribution

	genus, species, subspecies, continent, waterBody, islandGroup, island, countryCode, stateProvince, municipality, locality, habitat, lifeStage, behavior, establishmentMeans, organismQuantity, organismQuantityType, sex, reproductiveCondition, pathway, degreeOfEstablishment, associatedTaxa, organismRemarks, samplingProtocol, sampleSizeValue, sampleSizeUnit, samplingEffort, fieldNotes, eventRemarks, iucnRedListCategory	owner_location,



Supplementary Figure 5. Location uncertainty distributions across image repositories. a) Box plots showing the distribution of coordinate uncertainty (in meters, log scale) for iNaturalist (iNat) and GBIF observations. The boxes represent the interquartile range, with the horizontal line showing the median value. Whiskers extend to the most extreme non-outlier values, with outliers plotted as individual points. b): Distribution of Flickr's accuracy levels, ranging from 1 (world level) to 16 (street level). Red dashed lines and labels indicate key geographic reference points in Flickr's scale: world (1), country (3), region (6), city (11), and street level (16).

SUPPLEMENTARY MATERIAL, CHAPTER 5

For Taxonomic Classification and keywords of Salmonidae Family used to search photo-sharing sites, see Supplementary Table 6.

Image attribution for Figure 5.5:

a) Adapted (resized) from Photo 164294760 (https://www.inaturalist.org/photos/164294760) by Tom Clenche (https://www.inaturalist.org/people/tclenche), some rights reserved (CC BY-NC) [https://creativecommons.org/licenses/by-nc/4.0/]
b) Adapted (resized) from Photo 75576535 (https://www.inaturalist.org/photos/75576535) by anniezgrab (https://www.inaturalist.org/people/anniezgrab), some rights reserved (CC BY-NC) [https://creativecommons.org/licenses/by-nc/4.0/]
c) Adapted (resized) from Photo 105571877 (https://www.inaturalist.org/photos/105571877) by Liadan Dickie (https://www.inaturalist.org/people/fuligogirl), some rights reserved (CC BY-NC) [https://creativecommons.org/licenses/by-nc/4.0/]
d) Adapted (resized) from Photo 309595571 (https://www.inaturalist.org/photos/309595571) by moisearthur (https://www.inaturalist.org/people/moisearthur), some rights reserved (CC BY-NC) [https://creativecommons.org/licenses/by-nc/4.0/]
e) Adapted (resized and overlayed with heatmap) from Photo 164294760 (https://www.inaturalist.org/photos/164294760) by Tom Clenche (https://www.inaturalist.org/people/tclenche), some rights reserved (CC BY-NC) [https://creativecommons.org/licenses/by-nc/4.0/]
f) Adapted (resized and overlayed with heatmap) from Photo 75576535 (https://www.inaturalist.org/photos/75576535) by anniezgrab

(https://www.inaturalist.org/people/anniezgrab), some rights reserved (CC BY-NC) [https://creativecommons.org/licenses/by-nc/4.0/]

g) Adapted (resized and overlayed with heatmap) from Photo 105571877 (https://www.inaturalist.org/photos/105571877) by Liadan Dickie (https://www.inaturalist.org/people/fuligogirl), some rights reserved (CC BY-NC) [https://creativecommons.org/licenses/by-nc/4.0/]

h) Adapted (resized and overlayed with heatmap) from Photo 309595571 (https://www.inaturalist.org/photos/309595571) by moisearthur (https://www.inaturalist.org/people/moisearthur), some rights reserved (CC BY-NC) [https://creativecommons.org/licenses/by-nc/4.0/]

Image Quality Analysis

Methods

To quantitatively assess the impact of image quality on model performance, we calculated three no-reference image quality metrics for each image in the Salmo genus-specific validation dataset. We used the variance of the Laplacian as a proxy for image sharpness, where a higher score indicates a sharper image. We also calculated two perceptual quality scores: the Blind/Referenceless Image Spatial Quality Evaluator (BRISQUE) score (Mittal et al. 2012) and the Natural Image Quality Evaluator (NIQE) score (Mittal et al. 2013). For both BRISQUE and NIQE, a lower score indicates better perceptual quality. Image dimensions (width and height in pixels) were also recorded.

We compared the distributions of these metrics between correctly and incorrectly classified images for the best-performing model (EfficientNetV2S). The summary statistics are presented below. A Mann-Whitney U test was used to test for significant differences between the two groups (correct vs. incorrect) for each metric, with a significance level of α=0.05. Tests

were performed on the full, pooled dataset (all correct vs. all incorrect classifications) to assess the overall effect, and on a stratified dataset (correct vs. incorrect within each 'Healthy' and '*Saprolegnia* spp.' class) to investigate class-specific effects.

Results

A Mann-Whitney U test performed on the full, pooled dataset found no statistically significant differences between correctly and incorrectly classified images for any of the five metrics tested. The p-values for the pooled tests were: sharpness (p=0.366), BRISQUE (p=0.266), NIQE (p=0.787), image width (p=0.993), and image height (p=0.940).

A secondary, stratified analysis comparing correct and incorrect classifications within each of the 'Healthy' and 'Saprolegnia spp.' classes also found no statistically significant differences for any metric. While the summary statistics (Tables E1-E5) show minor differences in the means between groups, the statistical tests confirm these variations are not significant.

Supplementary Table 8. Sharpness (variance of Laplacian) statistics for images in the Salmo genus-specific data set, classified by the EfficientnetV2S model. Higher is better.

True Label	**Classification**	**count**	**mean**	**std**	**min**	**25%**	**50%**	**75%**	**max**
Healthy	Correct	956	670.74	1721.1	2.87	49.0	176.0	628.2	29270.93
Healthy	Incorrect	14	354.13	388.52	2.82	38.39	194.33	624.41	1268.82
Sapro	Correct	252	1439.79	2622.98	1.75	9.7	106.77	1479.77	15906.8
Sapro	Incorrect	3	82.28	82.76	4.86	38.67	72.48	120.99	169.5

Supplementary Table 9. BRISQUE score statistics for images in the Salmo genus-specific data set, classified by the EfficientnetV2S model. Lower is better.

True Label	Classification	count	mean	std	min	25%	50%	75%	max
Healthy	Correct	956	27.56	14.47	0.0	16.26	25.92	36.81	92.47
Healthy	Incorrect	14	33.12	16.87	8.32	17.93	33.14	46.99	60.21
Sapro	Correct	252	33.9	18.93	0.0	18.92	35.5	47.55	67.46
Sapro	Incorrect	3	38.87	37.76	0.0	20.6	41.2	58.31	75.42

Supplementary Table 10. NIQE score statistics for images in the Salmo genus-specific data set, classified by the EfficientnetV2S model. Lower is better.

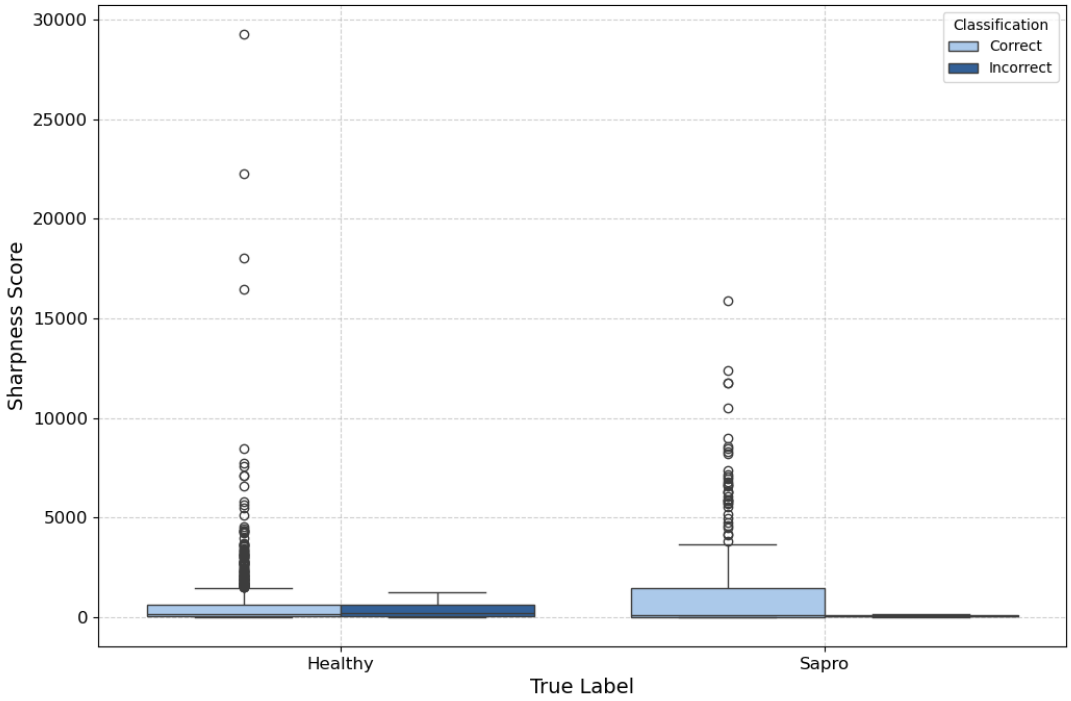
True Label	Classification	count	mean	std	min	25%	50%	75%	max
Healthy	Correct	956	14.45	4.63	4.57	11.67	13.77	16.44	69.33
Healthy	Incorrect	14	15.53	4.19	11.36	13.27	14.09	16.86	26.3
Sapro	Correct	252	15.13	5.49	6.54	11.31	14.67	17.5	50.07
Sapro	Incorrect	3	11.99	1.29	11.12	11.25	11.38	12.42	13.46

Supplementary Table 11. Height statistics for images in the Salmo genus-specific data set, classified by the EfficientnetV2S model.

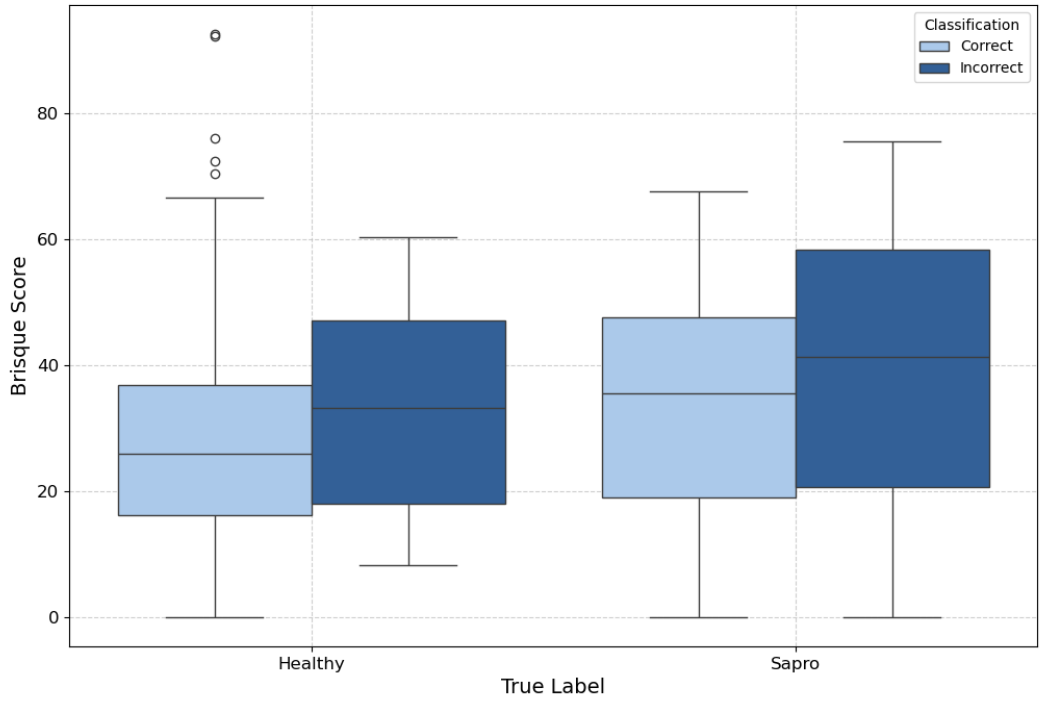
True Label	Classification	count	mean	std	min	25%	50%	75%	max
Healthy	Correct	956	1618.56	699.41	165.0	1160.5	1536.0	2048.0	5616.0
Healthy	Incorrect	14	1355.93	484.57	300.0	1152.0	1200.0	1554.75	2048.0
Sapro	Correct	252	1604.16	1203.91	159.0	579.0	1268.5	2592.0	5312.0
Sapro	Incorrect	3	3578.67	1524.15	2448.0	2712.0	2976.0	4144.0	5312.0

Supplementary Table 12. Width statistics for images in the Salmo genus-specific data set, classified by the EfficientnetV2S model.

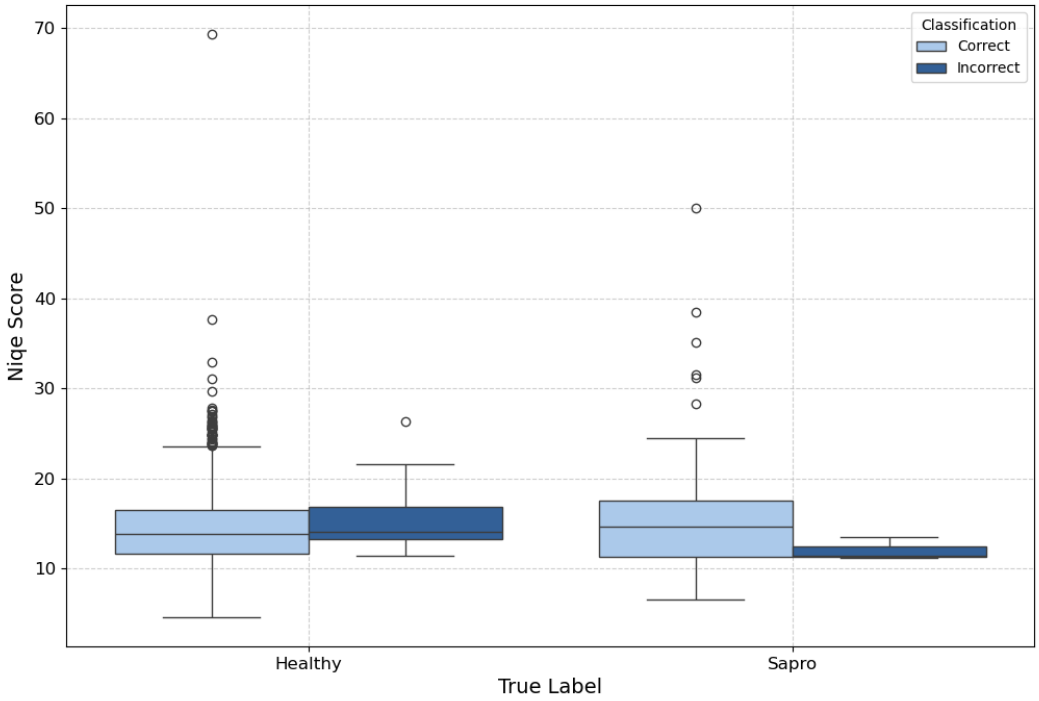
True Label	Classification	count	mean	std	min	25%	50%	75%	max
Healthy	Correct	956.0	1895.88	890.33	287.0	1536.0	2048.0	2048.0	6720.0
Healthy	Incorrect	14.0	1535.64	529.15	640.0	1164.0	1568.0	2048.0	2048.0
Sapro	Correct	252.0	2137.21	1780.69	213.0	480.0	1536.0	4032.0	5312.0
Sapro	Incorrect	3.0	3406.67	505.34	2988.0	3126.0	3264.0	3616.0	3968.0



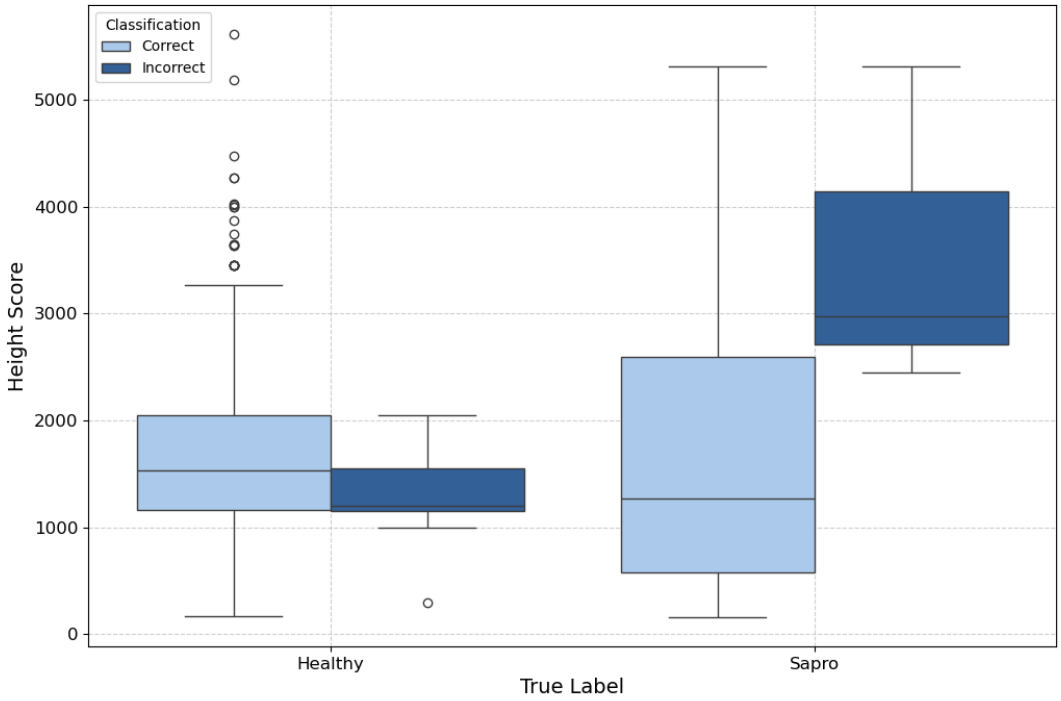
Supplementary Figure 6. Box plot showing median, interquartile range, minimum and maximum values of sharpness (variance of Laplacian) scores for images in the Salmo genus-specific data set, classified by the EfficientnetV2S model. Box plots show median, interquartile range, minimum and maximum values.



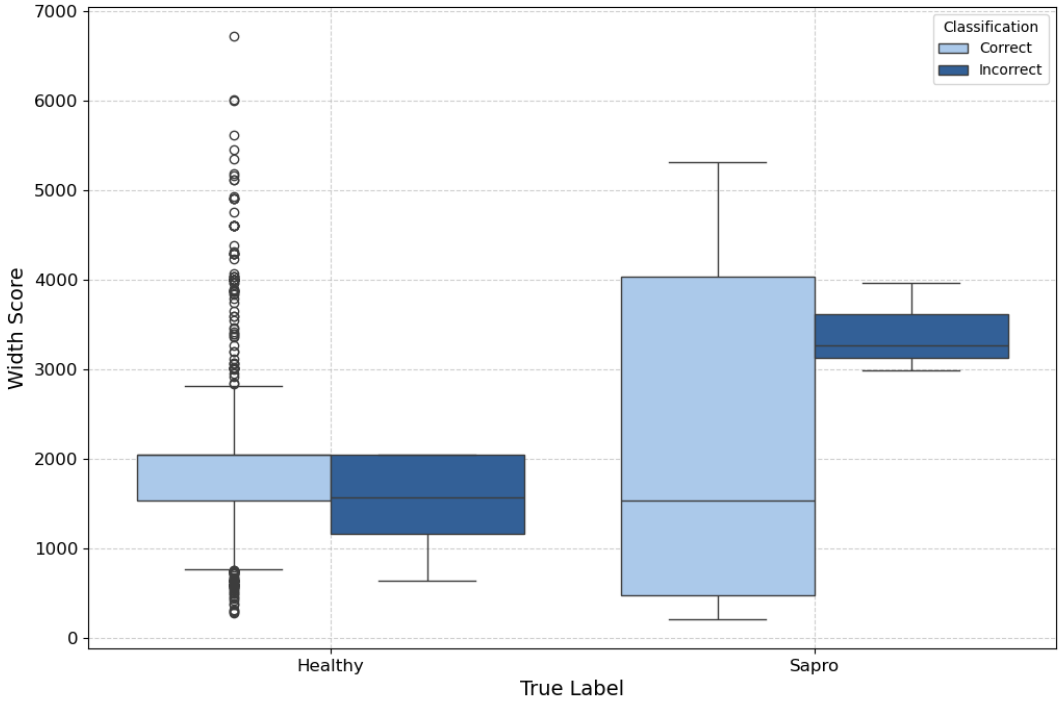
Supplementary Figure 7. Box plot showing median, interquartile range, minimum and maximum values of BRISQUE scores for images in the Salmo genus-specific data set, classified by the EfficientnetV2S model.



Supplementary Figure 8. Box plot showing median, interquartile range, minimum and maximum values of NIQE scores for images in the Salmo genus-specific data set, classified by the EfficientnetV2S model.

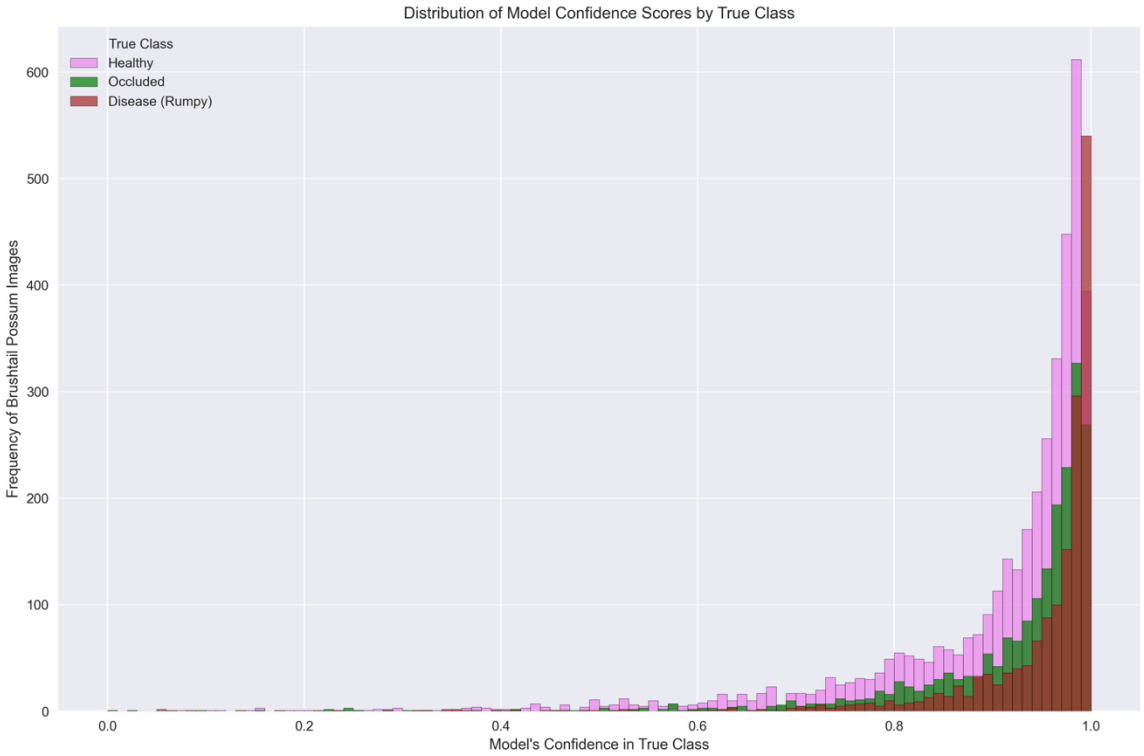


Supplementary Figure 9. Box plot showing median, interquartile range, minimum and maximum values of height for images in the Salmo genus-specific data set, classified by the EfficientnetV2S model.

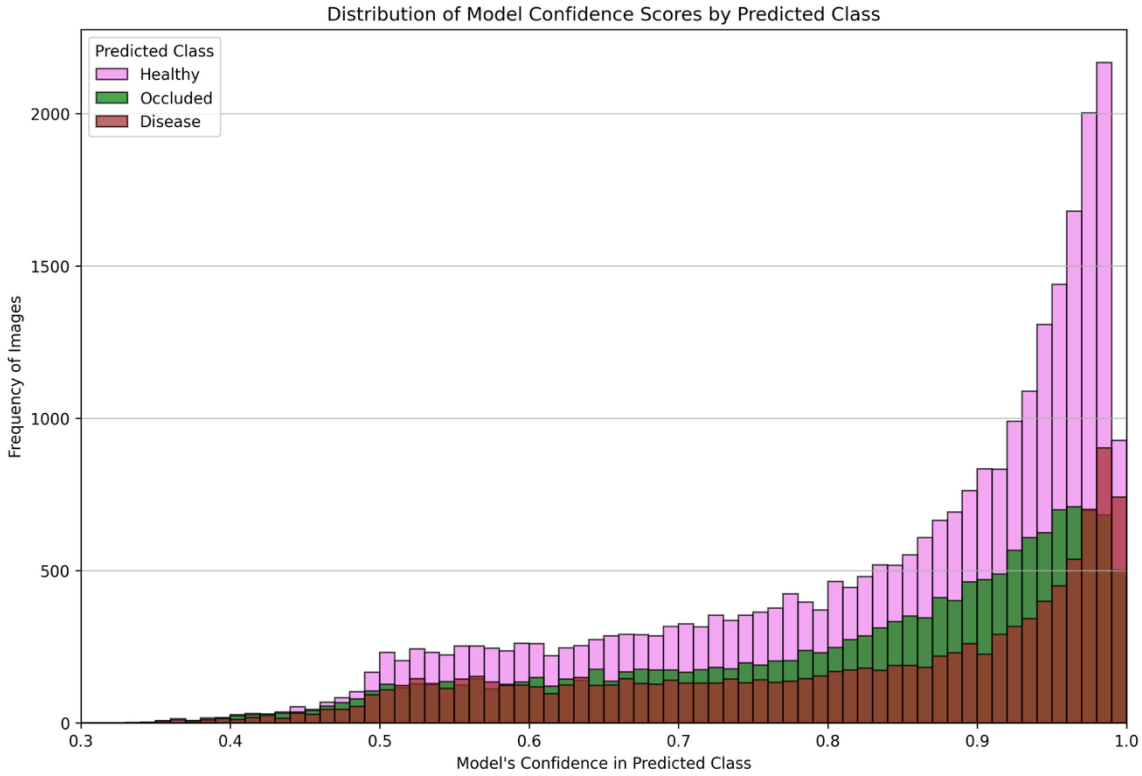


Supplementary Figure 10. Box plot showing median, interquartile range, minimum and maximum values of width for images in the Salmo genus-specific data set, classified by the EfficientnetV2S model.

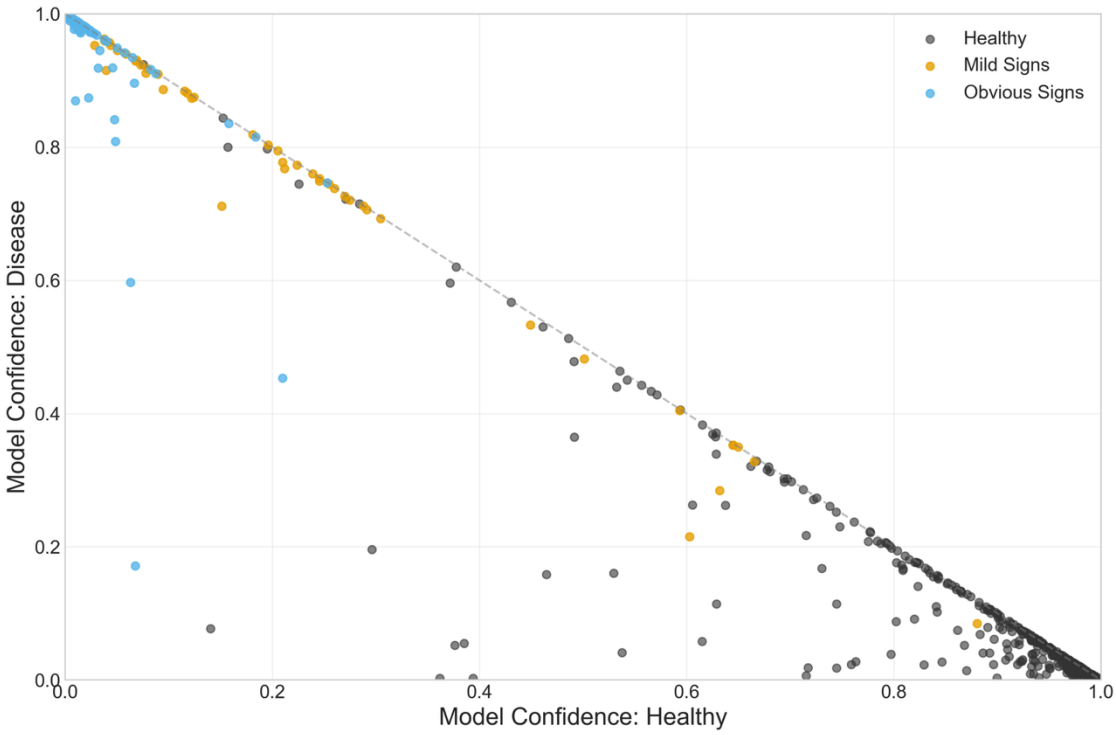
SUPPLEMENTARY MATERIAL, CHAPTER 6



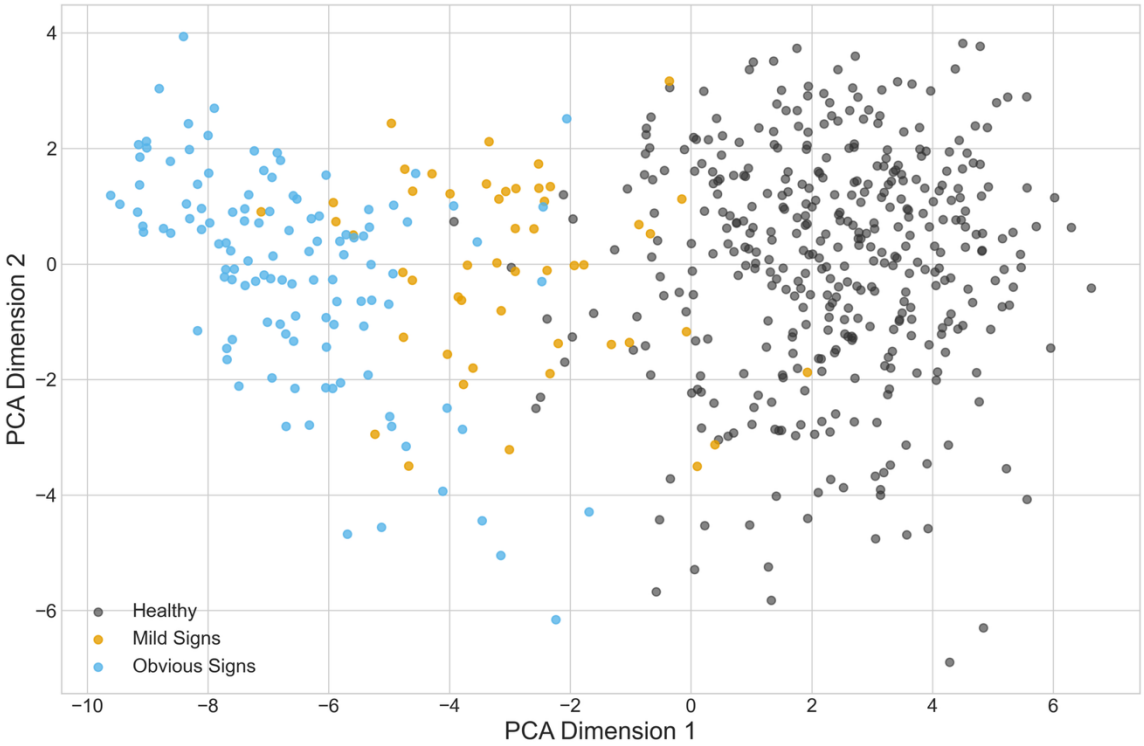
Supplementary Figure 11. Distribution of the model's confidence scores across training, validation and test images, categorised by ground truth class.



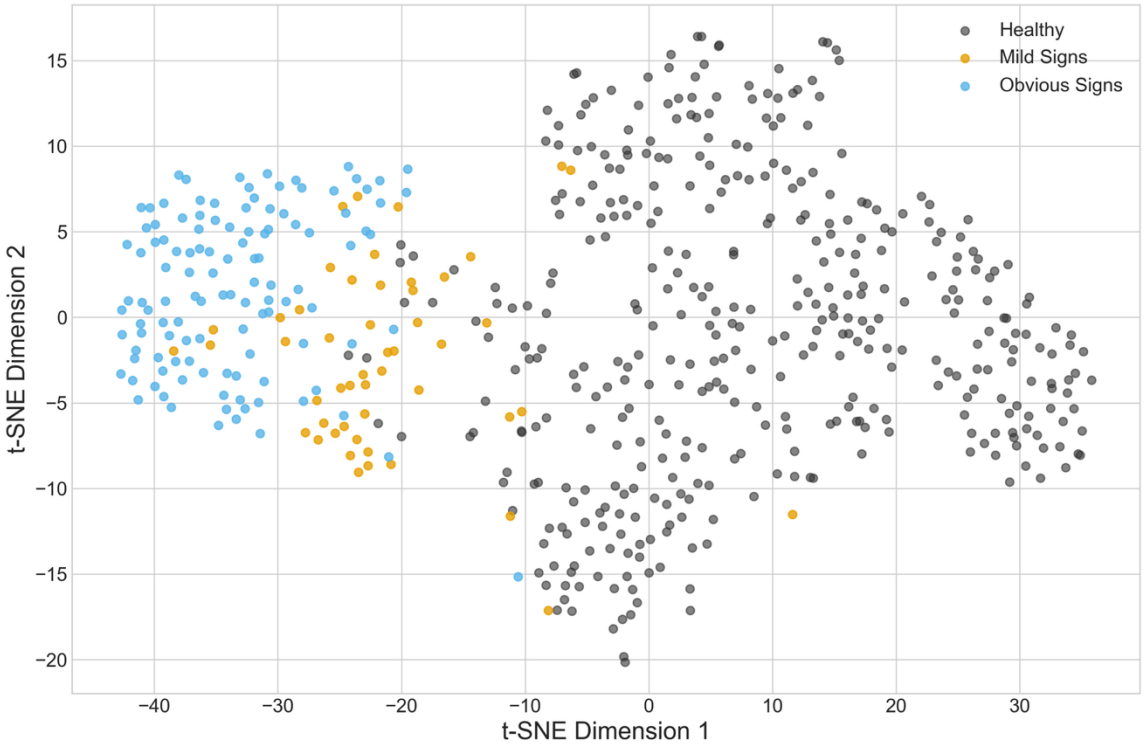
Supplementary Figure 12. Distribution of the model's confidence scores across all unlabelled images, categorised by the model's predicted class.



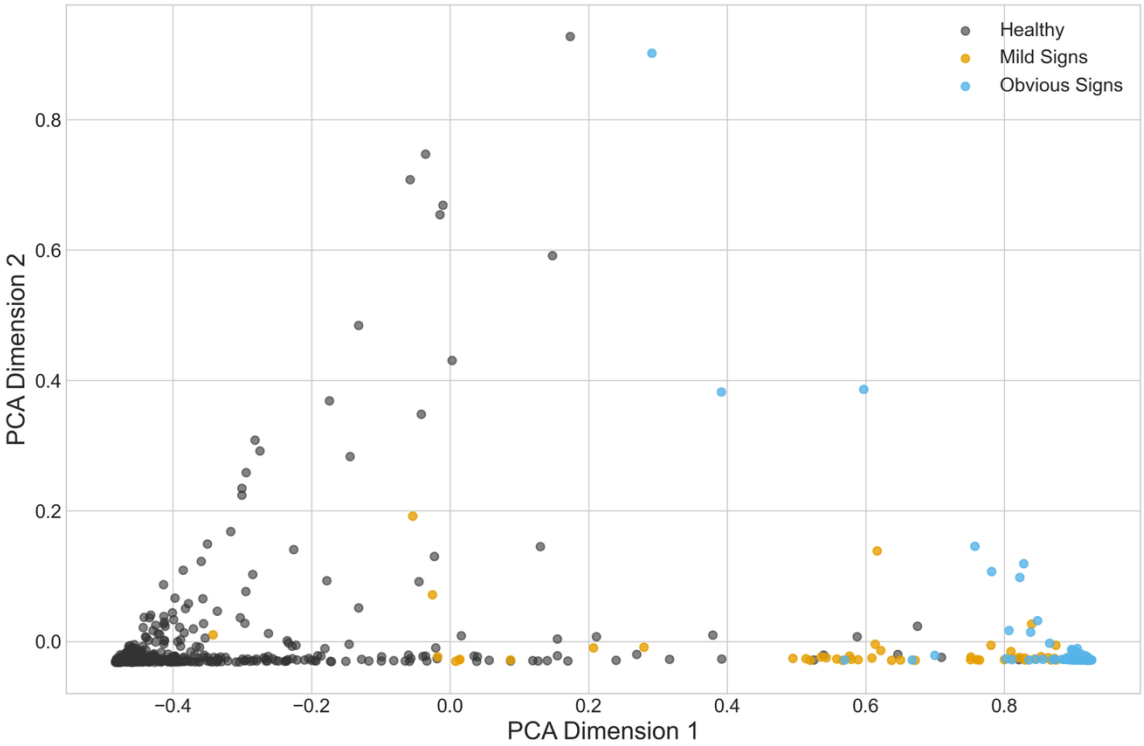
Supplementary Figure 13. Scatter plot of the supervised model's output probabilities for the 'Healthy' and 'Disease' classes on the non-occluded test set. Each point represents an image, coloured by its expert-assigned ground truth label.



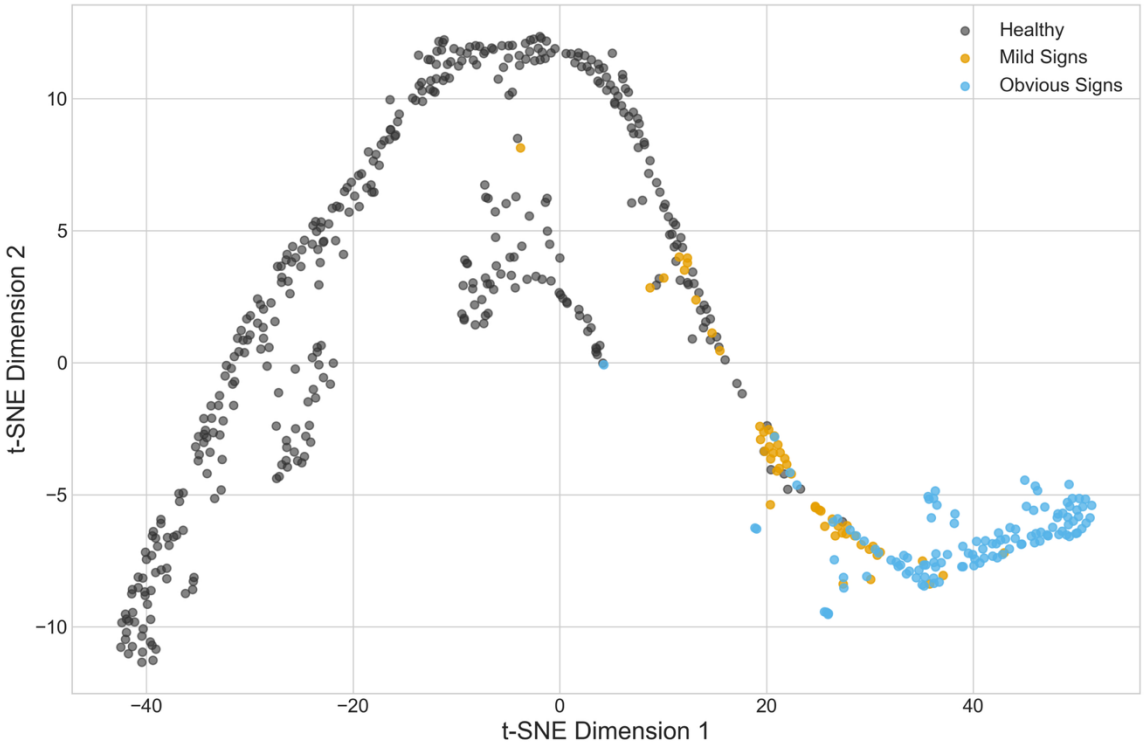
Supplementary Figure 14. Principal Component Analysis (PCA) of the bottleneck features extracted from the test set images. Each point represents an individual image, plotted according to the first two principal components.



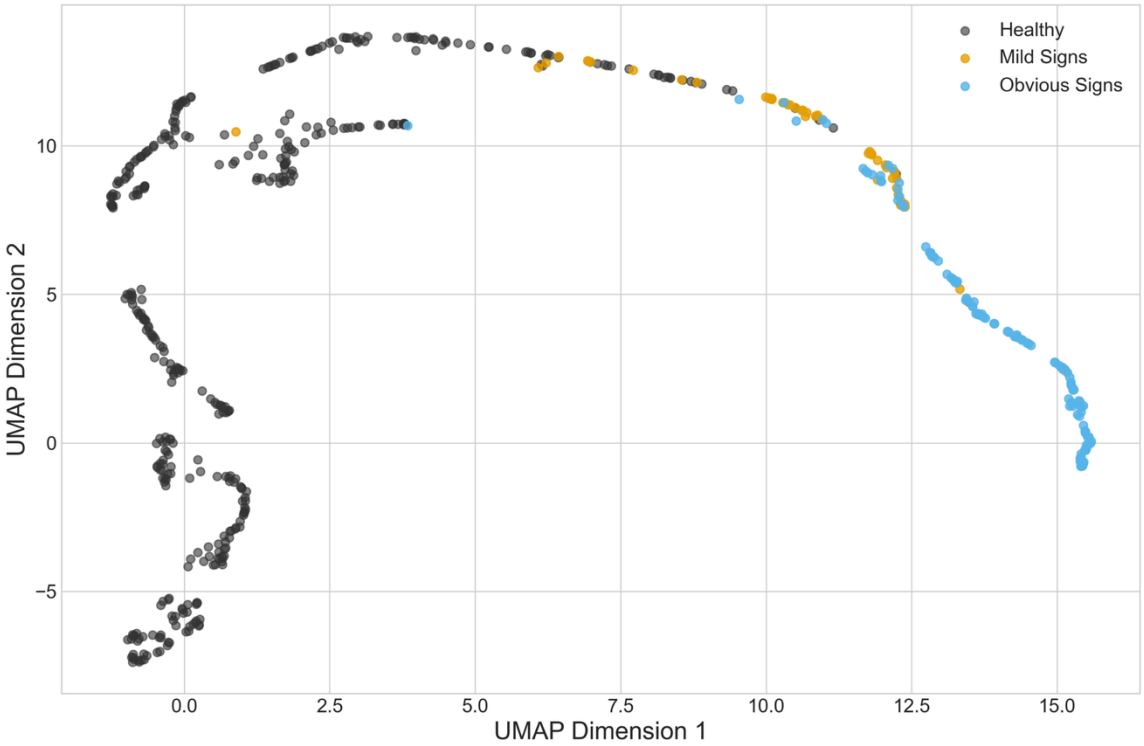
Supplementary Figure 15. t-Distributed Stochastic Neighbour Embedding (t-SNE) of the bottleneck features extracted from the test set images. Each point represents an individual image, plotted in the two-dimensional embedding space.



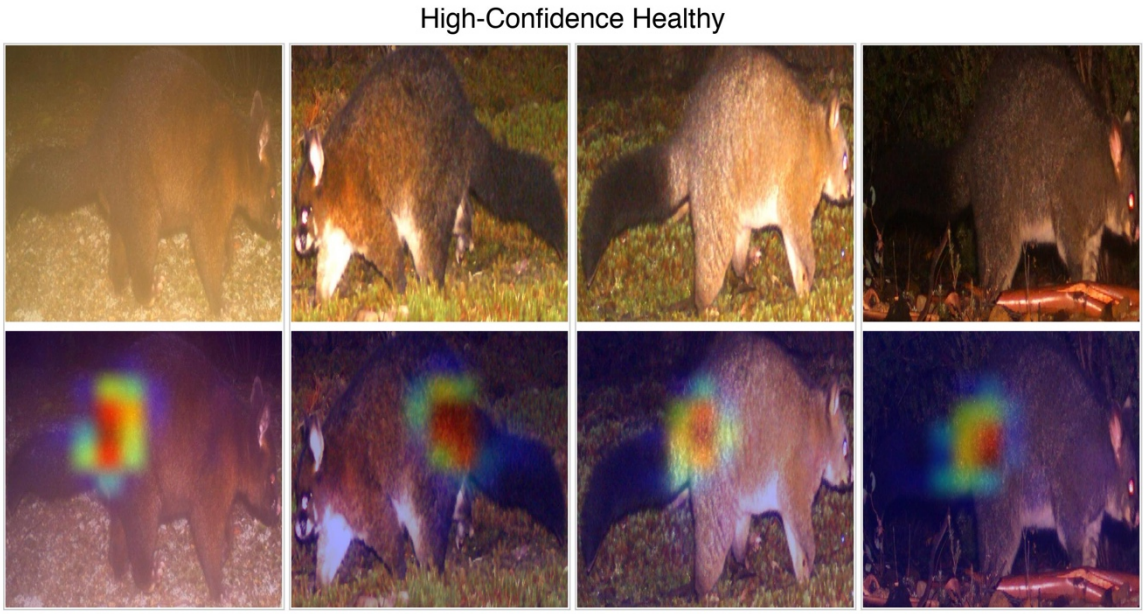
Supplementary Figure 16. Principal Component Analysis (PCA) of the final prediction probabilities from the test set. Each point represents an individual image, plotted according to the first two principal components derived from the model's three-class softmax outputs.



Supplementary Figure 17. t-Distributed Stochastic Neighbour Embedding (t-SNE) of the final prediction probabilities from the test set. Each point represents an individual image, plotted in the two-dimensional embedding space derived from the model's three-class softmax outputs.

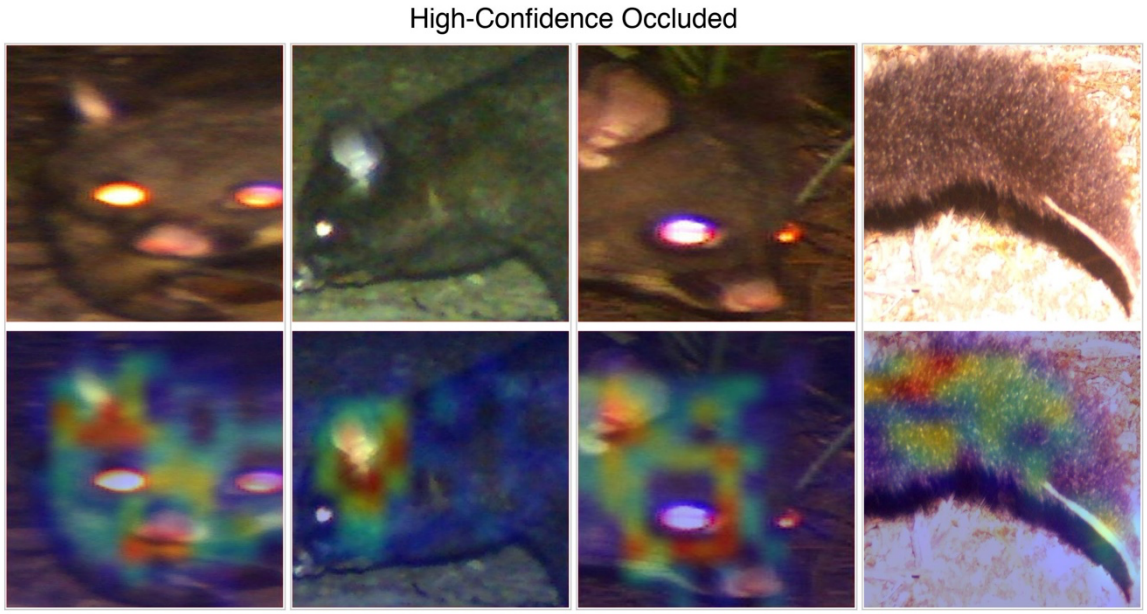


Supplementary Figure 18. Uniform Manifold Approximation and Projection (UMAP) of the final prediction probabilities from the test set. Each point represents an individual image, plotted in the two-dimensional embedding space derived from the model's three-class softmax outputs.

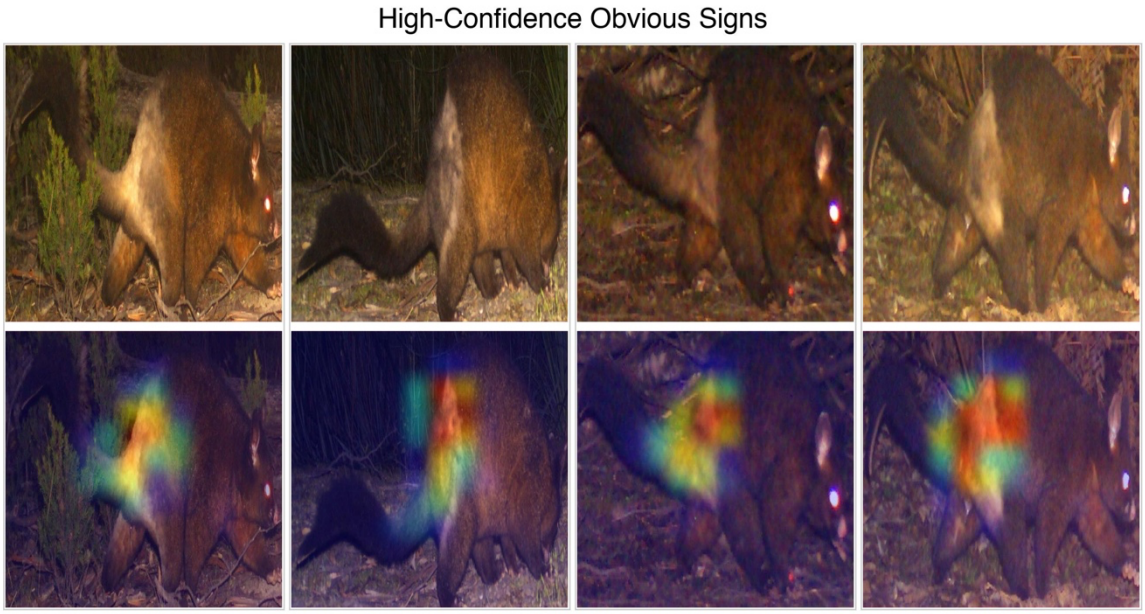


Supplementary Figure 19. Additional Grad-CAM visualisations for the high-confidence ‘Healthy’ classification scenario. Each of the four panels displays an example image from the test set (top) and the same image with the Grad-CAM heatmap overlay (bottom). The

heatmap indicates image regions of high importance for the model's final prediction, with red signifying high importance and blue signifying low importance.

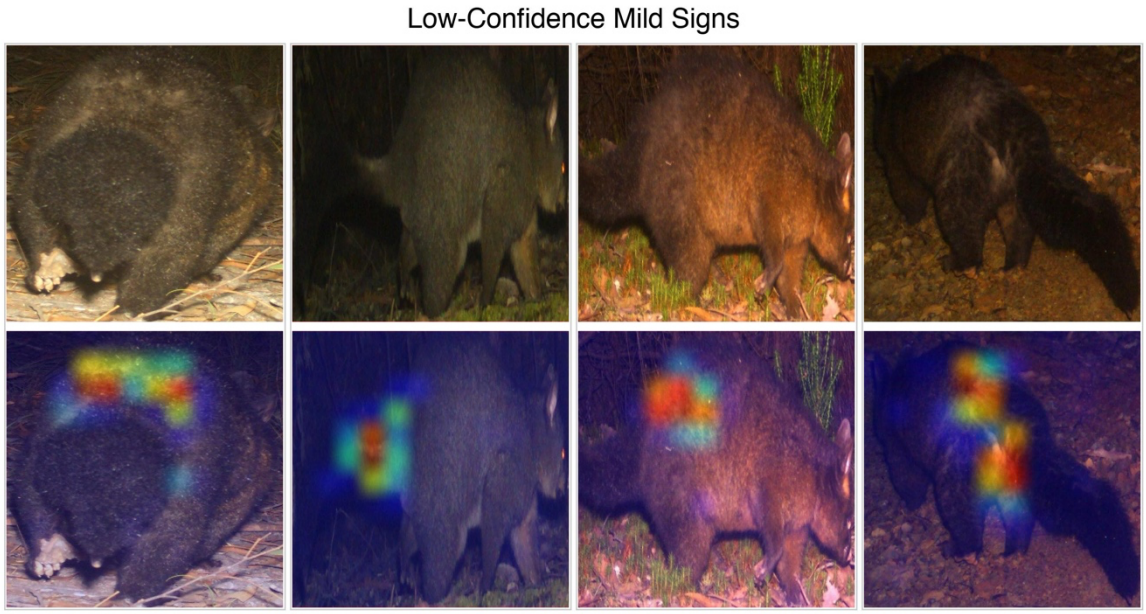


Supplementary Figure 20. Additional Grad-CAM visualisations for the high-confidence ‘Occluded’ classification scenario. Each of the four panels displays an example image from the test set (top) and the same image with the Grad-CAM heatmap overlay (bottom). The heatmap indicates image regions of high importance for the model's final prediction, with red signifying high importance and blue signifying low importance.

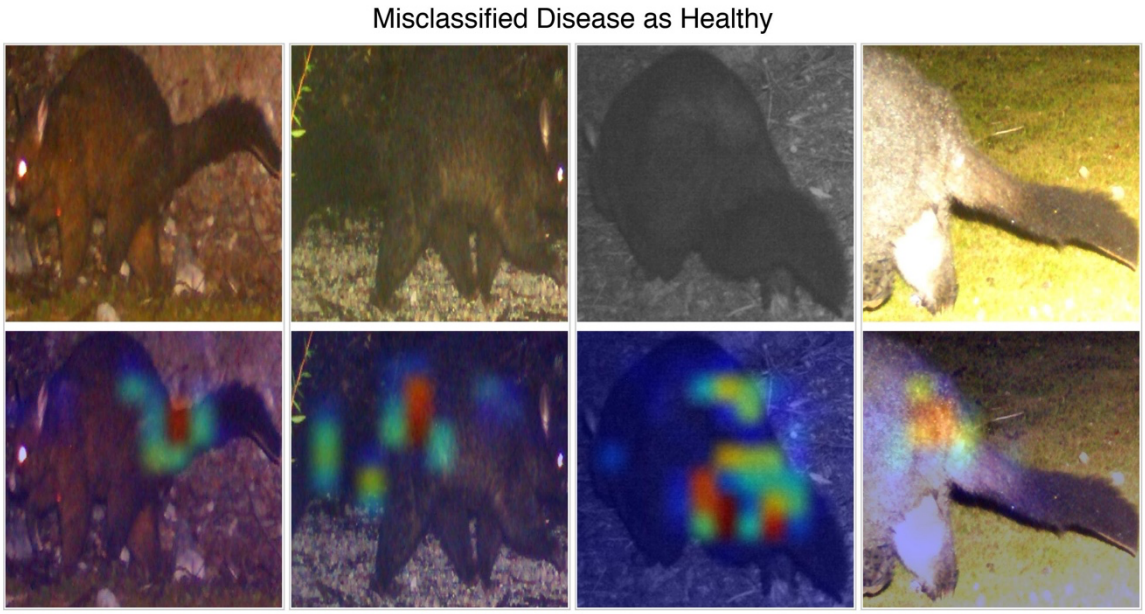


Supplementary Figure 21. Additional Grad-CAM visualisations for the high-confidence ‘Disease’ classification scenario, where the images belong to the ‘Obvious signs’ class. Each of the four panels displays an example image from the test set (top) and the same image with the Grad-CAM heatmap overlay (bottom). The heatmap indicates image regions of high

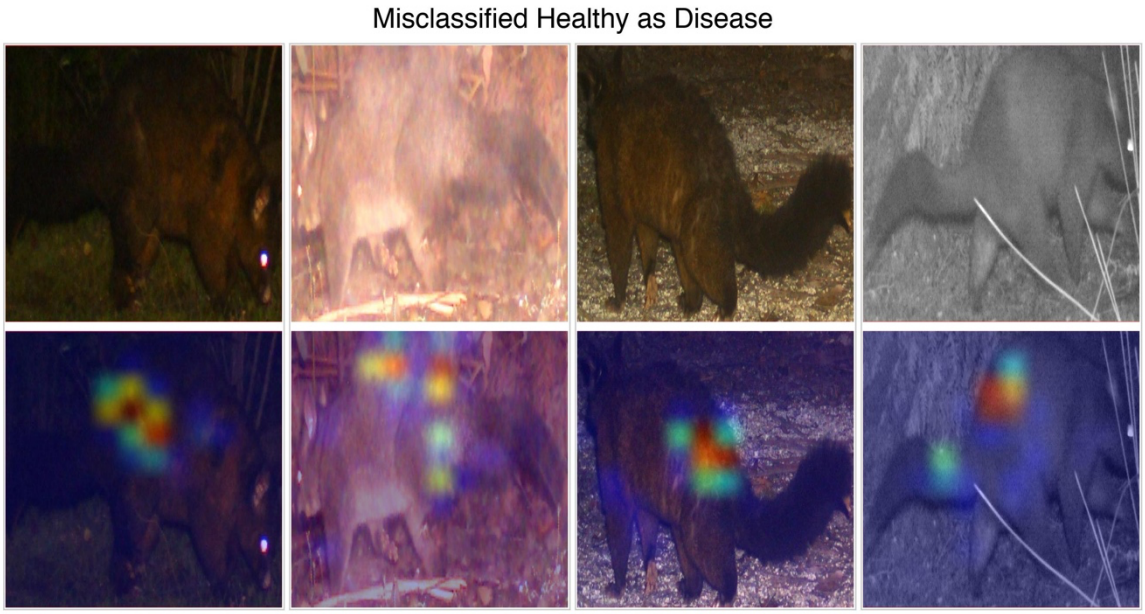
importance for the model's final prediction, with red signifying high importance and blue signifying low importance.



Supplementary Figure 22. Additional Grad-CAM visualisations for the low-confidence ‘Disease’ classification scenario, where the images belong to the ‘Mild signs’ class. Each of the four panels displays an example image from the test set (top) and the same image with the Grad-CAM heatmap overlay (bottom). The heatmap indicates image regions of high importance for the model's final prediction, with red signifying high importance and blue signifying low importance.



Supplementary Figure 23. Additional Grad-CAM visualisations for the divergent classifications of ‘Disease’ as ‘Healthy’ classification scenario. Each of the four panels displays an example image from the test set (top) and the same image with the Grad-CAM heatmap overlay (bottom). The heatmap indicates image regions of high importance for the model's final prediction, with red signifying high importance and blue signifying low importance.



Supplementary Figure 24. Additional Grad-CAM visualisations for the divergent classifications of ‘Healthy’ as ‘Disease’ classification scenario. Each of the four panels displays an example image from the test set (top) and the same image with the Grad-CAM heatmap overlay (bottom). The heatmap indicates image regions of high importance for the model's final prediction, with red signifying high importance and blue signifying low importance.

Supplementary Table 13. Test set performance comparison of the final semi-supervised learning model against its supervised baseline. Best values for each metric are shown in bold.

Metric	**Wildlife_ft**	**Wildlife_SSL**
Accuracy	**0.952**	0.948
F1-score (Healthy)	**0.958**	0.957
F1-score (Disease)	**0.926**	0.924

F1-score (Occluded)	**0.96**	0.95
Macro avg F1-score	**0.948**	0.944
Macro avg Precision	**0.947**	0.942
Macro avg Recall	**0.95**	0.946
Precision (Healthy)	**0.962**	0.959
Precision (Disease)	**0.913**	0.898
Precision (Occluded)	0.965	**0.969**
Recall (Healthy)	**0.955**	**0.955**
Recall (Disease)	0.94	**0.952**
Recall (Occluded)	**0.956**	0.931
Weighted avg F1-score	**0.952**	0.948
Weighted avg Precision	**0.952**	0.949
Weighted avg Recall	**0.952**	0.948

SUPPLEMENTARY MATERIAL, REFERENCES

Dimson, M. and Gillespie, T.W. 2023. Who, where, when: Observer behavior influences spatial and temporal patterns of iNaturalist participation. *Applied Geography* 153, p. 102916. doi: 10.1016/j.apgeog.2023.102916.

Lenth, R.V. 2024. emmeans: Estimated Marginal Means, aka Least-Squares Means. Available at: https://CRAN.R-project.org/package=emmeans.

Lüdecke, D., Ben-Shachar, M.S., Patil, I., Waggoner, P. and Makowski, D. 2021. performance: An R package for assessment, comparison and testing of statistical models. *Journal of open source software* 6(60).

Mittal, A., Moorthy, A.K. and Bovik, A.C. 2012. No-Reference Image Quality Assessment in the Spatial Domain. *IEEE Transactions on Image Processing* 21(12), pp. 4695–4708. doi: 10.1109/TIP.2012.2214050.

Mittal, A., Soundararajan, R. and Bovik, A.C. 2013. Making a “Completely Blind” Image Quality Analyzer. *IEEE Signal Processing Letters* 20(3), pp. 209–212. doi: 10.1109/LSP.2012.2227726.

Qiao, H., Orr, M., Yang, Q., Zhan, X., Lei, F. and Hughes, A.C. 2023. Global birdwatching data reveal uneven consequences of the COVID-19 pandemic. *Biological Conservation* 288, p. 110351. doi: 10.1016/j.biocon.2023.110351.

Sánchez-Clavijo, L.M. et al. 2021. Differential reporting of biodiversity in two citizen science platforms during COVID-19 lockdown in Colombia. *Biological Conservation* 256, p. 109077. doi: 10.1016/j.biocon.2021.109077.

Wooldridge, J.M. 2020. *Introductory econometrics: a modern approach*. Seventh edition. Australia: Cengage Learning.