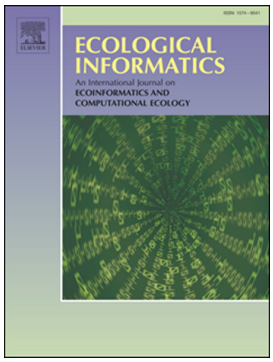


Computer vision for infectious disease surveillance; Saprolegnia spp. in salmonids

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

PII: S1574-9541(25)00576-X

DOI: https://doi.org/10.1016/j.ecoinf.2025.103567

Reference: ECOINF 103567

To appear in: *Ecological Informatics*

Received date: 29 April 2025

Revised date: 14 December 2025

Accepted date: 15 December 2025

Please cite this article as: A.S. Olsen, P.L. Rosin, C.B. Jones, et al., Computer vision for infectious disease surveillance; Saprolegnia spp. in salmonids, *Ecological Informatics* (2024), https://doi.org/10.1016/j.ecoinf.2025.103567

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

Corresponding author: Agnethe S. Olsen

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.

CRediT Author Statement

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.

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/sapro_classification_comparison*. 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 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.

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. We developed and evaluated a computer vision pipeline to detect saprolegniasis-like infections, a devastating disease in salmonids that manifests as visible signs.

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* spp. specific dataset, with a mean recall (proportion of infected fish images correctly identified) of 0.898 ($\pm$ 0.043) and precision (proportion of correctly identified infected fish among all fish identified as infected) of 0.858 ($\pm$ 0.067). Performance varied with host taxonomic scope, with models achieving lower metrics on broader host taxa datasets. 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.

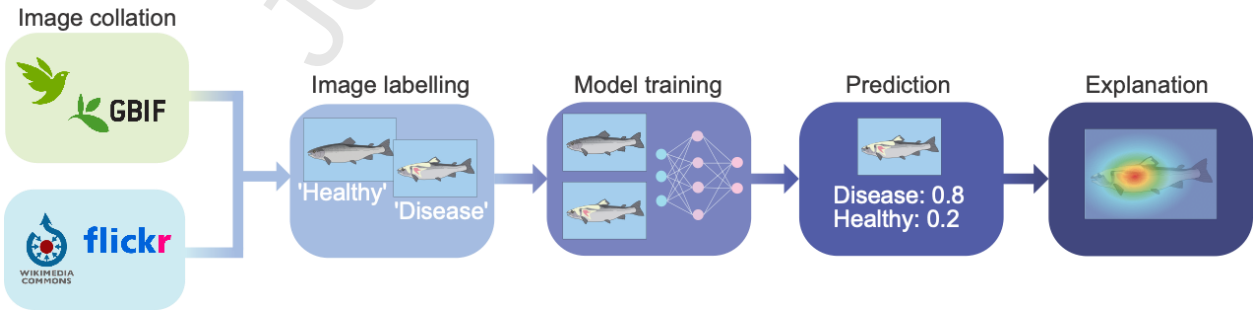
Computer vision-based surveillance could enable earlier outbreak detection and targeted diagnostics, improving freshwater ecosystem health management. While successful implementation hinges on acquiring sufficient high-quality imagery, this study highlights the potential of applying tailored Artificial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

Keywords: machine learning, disease ecology, iEcology, wildlife, surveillance, visible signs

Highlights

1. Computer vision pipeline detects *Saprolegnia* spp. disease signs in wild salmonids
2. Citizen science images viable source for fish disease detection
3. EfficientNetV2S achieves highest recall and precision for on *Salmo* spp. specific dataset
4. Model performance sensitive to host taxonomic level and specific image artefacts
5. AI enables large-scale non-invasive surveillance of wildlife disease

Visual Abstract:



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 in the field 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, characterized 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). *Ad hoc* reporting often constitutes one-off, unstructured, non-systematic alerts from anglers or members of the public who happen to spot diseased fish. Once an outbreak is recognised, 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 assemble a novel dataset for this disease from diverse online and stakeholder sources using images of wild-caught fish, and systematically assess classifier performance across different host taxonomies. This approach directly tests the utility of using existing online data as a scalable, cost-effective alternative to dedicated field surveys for disease surveillance. 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.

2 Methods

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.

2.1.1 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 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

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

(*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 Table S1 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 Table S1 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 Table S1. 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.

2.1.2 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 (Table S1) 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 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 for visible signs of disease, outlined in Figure 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 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.

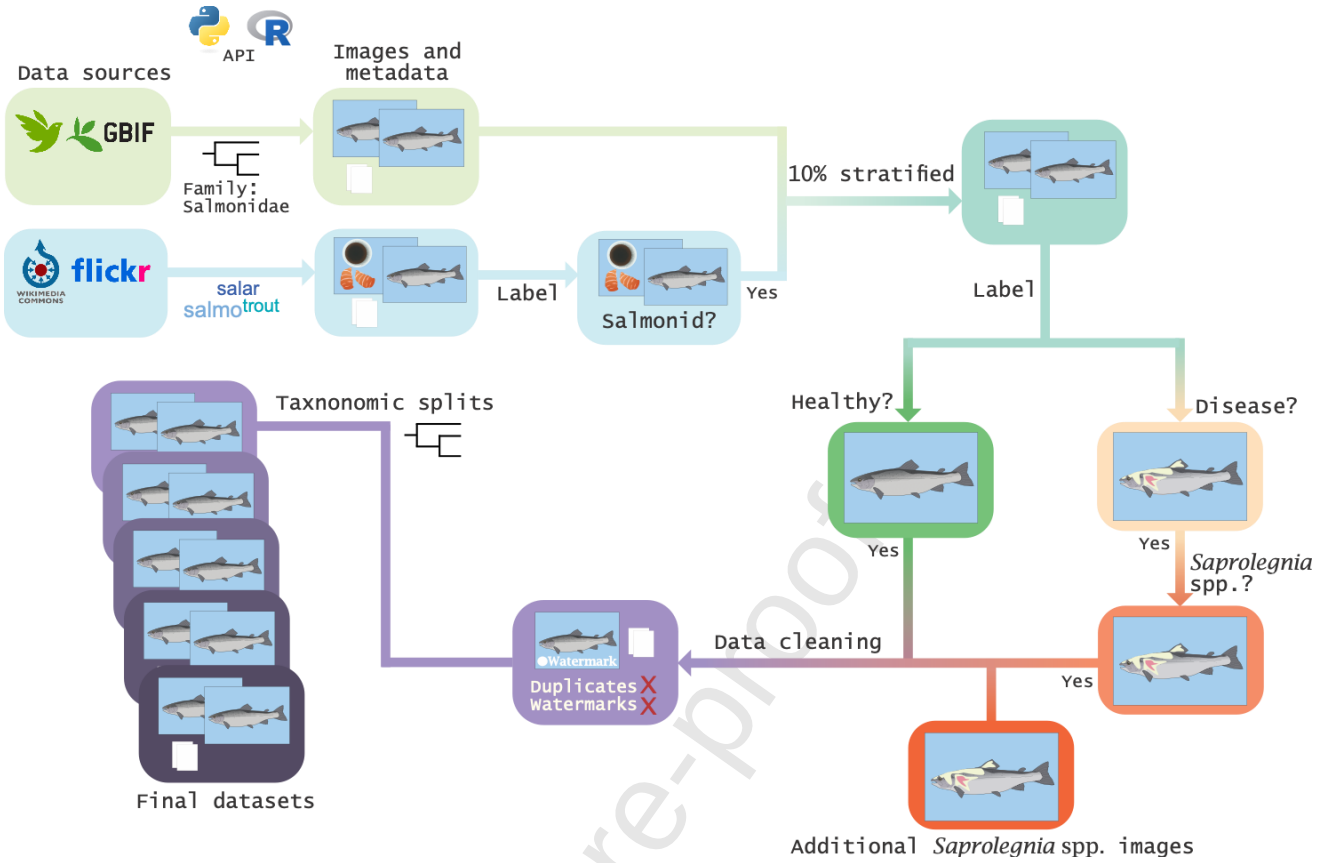


*Figure 1. Data pipeline. Data pipeline for collating images and identifying disease (*Saprolegnia *spp.) in salmonids. Images and metadata were downloaded from online sources using API searches based on taxonomy or search terms and false positives removed. A 10% subset, stratified by data source and 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.*

2.1.3 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 1).

Table 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

2.1.4 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.

2.1.5 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 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 1. 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 Table S1 for a full list of taxa.*

Host taxa	**Healthy**	***Saprolegnia* spp.**	**Ratio**
Salmo spp.	43	135	3.140
Salmo salar	91	110	1.209
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 2):

- 'Taxa present in both classes': This dataset included all images from taxa with images present in both the 'healthy' and '*Saprolegnia* spp.' classes (Table 1).
- '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 2). 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 2. 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

2.1.6 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 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.

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.

2.2.1 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. This will lead to poor generalisation, where the model performs poorly on new data, particularly for the minority class. The following sections detail the specific strategies we used to manage this challenge.

2.2.2 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. 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 generalisation capabilities. RandAugment, a data augmentation technique that applies a combination of image transformations (Cubuk et al. 2020), was applied to the oversampled data. RandAugment was implemented in Keras-CV with the number of augmentations, N, set to the default value of 3, and the magnitude, M, was set to the default value of 0.5.

2.2.3 Transfer learning

To leverage transfer learning, all models were instantiated with pre-trained weights from training on ImageNet (Krizhevsky et al. 2012). Models pre-trained on larger datasets, such as ImageNet, transfer the knowledge gained from training on large datasets to new tasks, such as classifying photographs of salmonids based on health status. The model can then be fine-tuned on this new, often smaller dataset.

2.2.4 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 $b = \log\left(\frac{1-\pi}{\pi}\right)$, 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.

2.2.5 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 (Fig. 2). Each cropped salmonid image was resized to the required input resolution (224x224 pixels for ResNet-50, MobileNetV3-S and EfficientNetV2-B0, 300x300 pixels for EfficientNetV2-S) 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 ResNet-50 and MobileNetV3-S, 256 for EfficientNetV2-B0 and 512 for EfficientNetV2-S), 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.

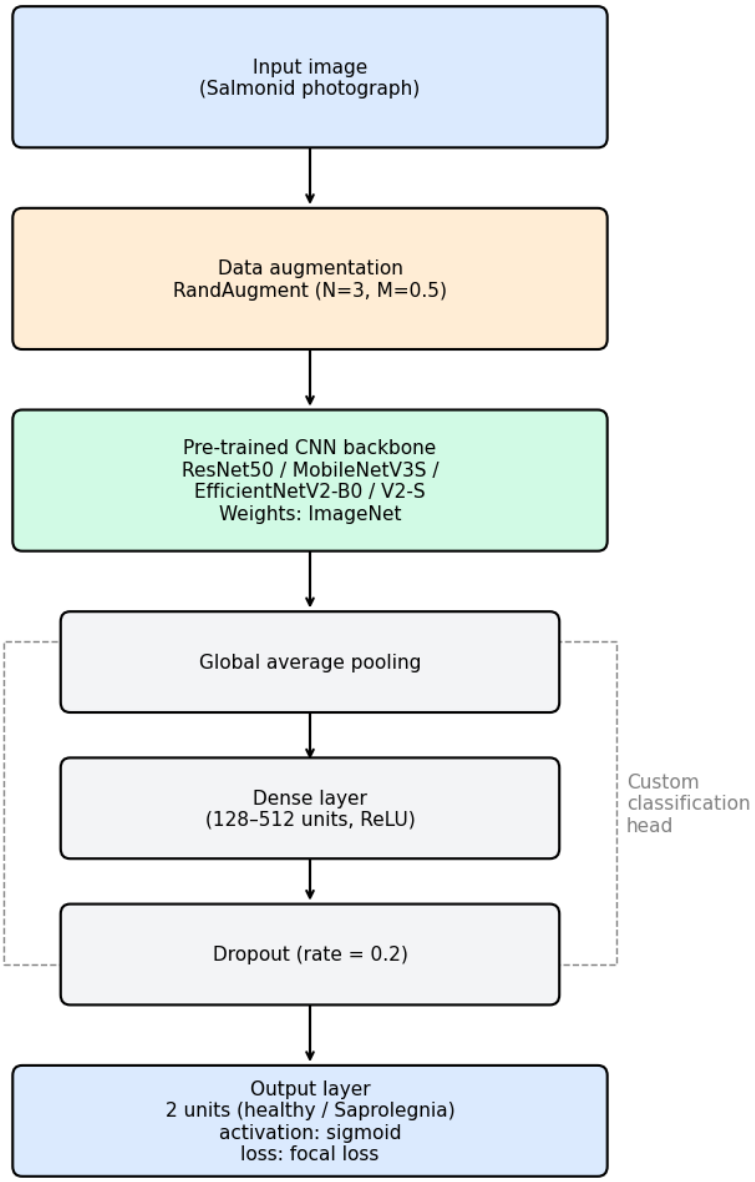


Figure 2. Transfer-learning pipeline used for training models on binary classification task. Each cropped salmonid photograph was resized and augmented using RandAugment (N = 3, M = 0.5) before being passed through a pre-trained convolutional backbone (ResNet-50, MobileNetV3-S, EfficientNetV2-B0 or EfficientNetV2-S, 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 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 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),t 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 parameters	Inference time per image [ms]

ResNet50	224×224	128	98	23850242	15216002	104.35
MobileNet V3Small	224×224	128	52	1013234	803114	27.38
EfficientNet V2B0	224×224	256	198	6247762	5707446	105.05
EfficientNet V2S	300×300	512	363	20988258	18876042	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.

2.2.6 Model Evaluation

We assessed models using common metrics (Géron 2022): precision (Eq. 1), recall (Eq. 2) and F1 (Eq. 3). We also calculated the Matthews Correlation Coefficient (MCC) (Eq. 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'.

$$Precision = \frac{TP}{TP + FP} \quad (1)$$

$$Recall = \frac{TP}{TP + FN} \quad (2)$$

$$F1 = 2 \times \frac{precision \times recall}{precision + recall} \quad (3)$$

$$MCC = 2 \times \frac{TP \times TN - FP \times FN}{\sqrt{(\mathrm{TP} + \mathrm{FP})(\mathrm{TP} + \mathrm{FN})(\mathrm{TN} + \mathrm{FP})(\mathrm{TN} + \mathrm{FN})}} \quad (4)$$

Precision (Eq. 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. 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. 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 Matthews Correlation Coefficient (MCC) (Eq. 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). In addition, we report the MCC, which provides a single summary value per model and dataset. From here on, we refer to class-wise metrics as metric_{class}, where 'metric' is one of the metrics precision, recall or F1, and 'class' is either 'healthy' or 'sapro' for *Saprolegnia* spp.

To further compare generalist and genus-specific training regimes on identical images, we carried out an additional exploratory evaluation in which all final models were applied to fixed Salmo and Oncorhynchus data sets. For each architecture and training dataset we computed $F1_{sapro}$ and MCC across the five cross-validation folds and summarised these as mean ± standard deviation. Full methodological details and results are provided in Supplementary Section S4.

2.2.7 Comparison to a random model

To establish a simple, unbiased benchmark, we employed a model that randomly assigned either the 'healthy' or '*Saprolegnia* spp.' class with equal probability and ran it 100 times to obtain a mean ratio. To compare the performance of our models against the random model, we used generalised linear models (GLM) using R (v4.1.2) with precision_{sapro}, recall_{sapro} and F1 statistics in turn as the response variables with model type (random baseline, ResNet50,

MobileNetV3S, EffienctNetV2B0, EfficientNetV2S) and data class as explanatory variables using a Gaussian error distribution.

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 Section S5.

3 Results

3.1 Model comparison

The EfficientNetV2S architecture demonstrated the best performance, yielding the highest values for all metrics (Fig. 3). EfficientNetV2B0, a smaller model requiring fewer computational resources, consistently achieved second-best performance across most

metrics. Interestingly, the maximum values were not achieved on the 'All images' but on the *Salmo* genus-specific dataset where EfficientNetV2S attained a macro-average F1 of 0.920 $\pm$ 0.029, recall$_{sapro}$ of 0.898 $\pm$ 0.043 and precision$_{sapro}$ of 0.858 $\pm$ 0.067. This high performance was likely due to the more balanced nature of the *Salmo* dataset, which contained a high proportion of '*Saprolegnia* spp.' class images from fishery stakeholders. Indeed, all the best metric scores were achieved on this taxonomic grouping, regardless of the classification model used (Fig. 3).

In contrast, the models generally performed worst on the *Oncorhynchus* genus-specific dataset, likely due to lower proportion of *Saprolegnia* spp. cases (Table 3) and greater visual variation between host taxa, particularly during spawning. MobileNetV3S produced the lowest precision$_{sapro}$ of 0.304 $\pm$ 0.086, F1$_{sapro}$ of 0.406 $\pm$ 0.081 and macro-average F1 of 0.640 $\pm$ 0.054 on the *Oncorhynchus* genus-specific dataset, while ResNet50 achieved the lowest recall$_{sapro}$ of 0.530 $\pm$ 0.080 on the 'Taxa with ≥ 10 photographs in both classes' dataset. However, even these lower-performing models still significantly exceeded random performance ($p < 5E^{-5}$) for recall$_{sapro}$, precision$_{sapro}$, and macro-average F1-score, particularly on the *Salmo* genus-specific dataset (Fig. 3).

All model architectures performed well on the 'Healthy' class metrics for all datasets. However, the models achieved their best performance, in terms of '*Saprolegnia* spp.' class-specific metrics and macro-average F1, on the *Salmo* dataset, followed by the broader datasets and again by the *Oncorhynchus* dataset. This pattern of declining performance across datasets was consistent across all models except for ResNet50, whose anomalous performance on the 'Taxa with ≥ 10 photographs in both classes' dataset (with the lowest recall$_{sapro}$) represented a deviation from the otherwise consistent trend. The Matthews

Correlation Coefficient showed the same ordering of models and datasets as the F1-based metrics, with EfficientNetV2S and EfficientNetV2B0 giving the highest values on all datasets (Fig. S1). This supports the use of F1 and macro-F1 as primary summary measures, as the conclusions are unchanged when using a metric that is less sensitive to class imbalance. The trend was especially evident for the $\text{precision}_{\text{sapro}}$ score of EfficientNetV2S, which decreased from 0.858 ± 0.067 on the *Salmo*-specific dataset to 0.656 ± 0.039 on the 'Taxa with ≥ 10 photographs in both classes' dataset, and further to 0.462 ± 0.041 on the *Oncorhynchus*-specific dataset (Fig. 3a).

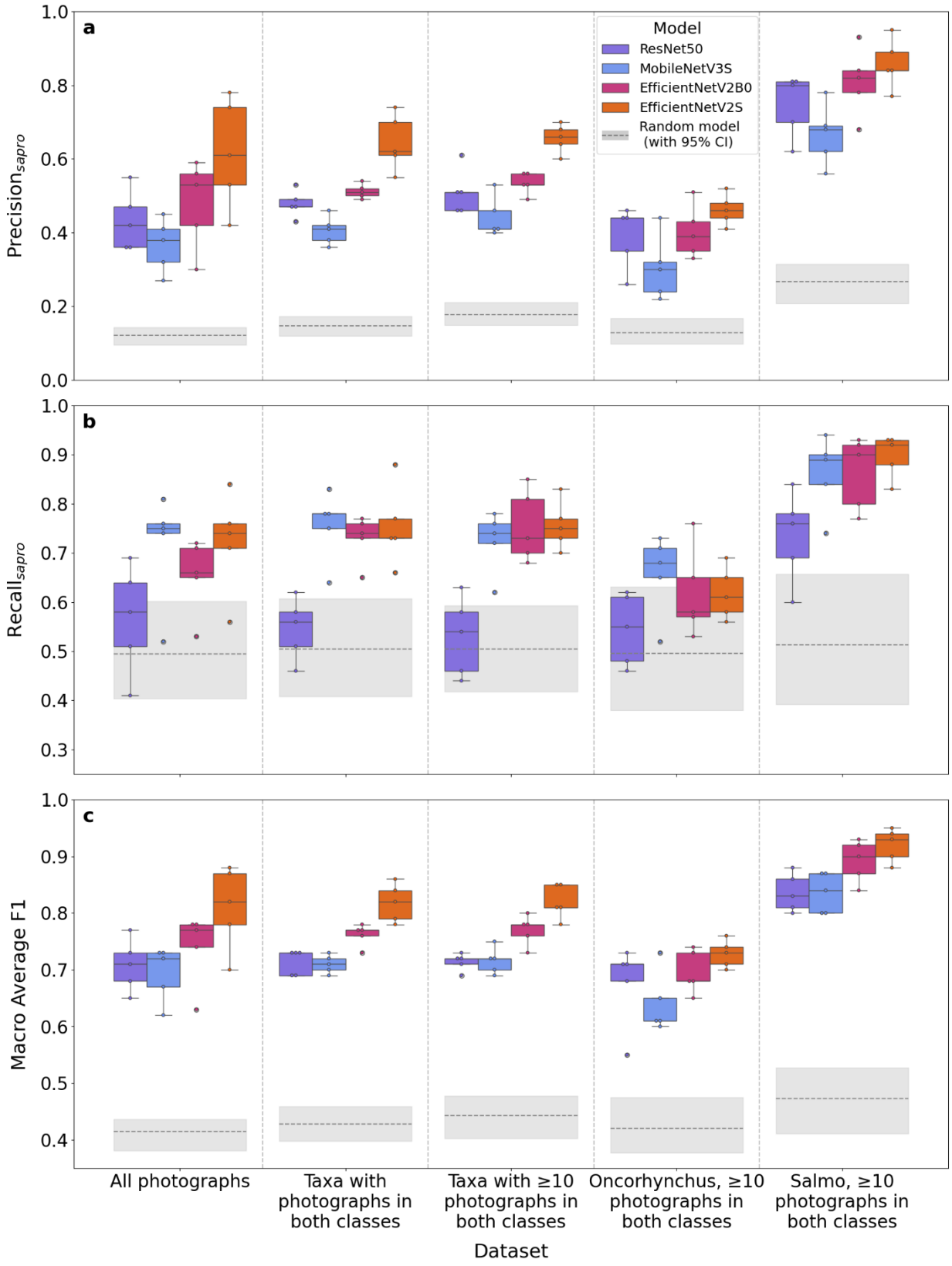


Figure 3. Performance metrics. Comparisons of precision$_{sapro}$, recall$_{sapro}$ and macro-average F1 for different models across the different tiered datasets. A random model with 95%

*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, minimum and maximum values, with individual points for each model run overlaid.*

While EfficientNetV2S generally showed strong performance, the model exhibited the highest variability in $precision_{sapro}$ and macro-average F1 on the 'All photographs' dataset (Fig. 3). In contrast, MobileNetV3S yielded the highest mean $recall_{sapro}$ for both the 'Taxa with photographs in both classes' and the *Oncorhynchus*-specific datasets (Fig. 3b), despite having the lowest $precision_{sapro}$ scores (Fig. 3a).

Across all datasets, MCC showed the same ordering and trends as the class-specific F1 and macro-average F1 scores (Fig. S1). On the '*Salmo*, ≥10 photographs in both classes' dataset, EfficientNetV2-S had the highest mean MCC (about 0.85), followed by EfficientNetV2-B0 (about 0.79), MobileNetV3-S (about 0.69) and ResNet-50 (about 0.67). For the broader datasets ('All photographs', 'Taxa with photographs in both classes' and 'Taxa with ≥10 photographs in both classes'), mean MCC values across models were lower, generally between 0.4 and 0.75, with the same ranking of architectures. On the '*Oncorhynchus*, ≥10 photographs in both classes' dataset, MCC values were lowest overall and ranged from roughly 0.23–0.35 for MobileNetV3-S and 0.39–0.46 for ResNet-50 to about 0.38–0.61 for EfficientNetV2-B0 and 0.53–0.71 for EfficientNetV2-S.

The exploratory analysis of applying all models to the *Salmo* and *Oncorhynchus* genus-specific datasets (Table S2) showed that models trained on the matching genus achieved the highest $F1_{sapro}$ and MCC. For example, EfficientNetV2-S trained on the *Salmo* dataset reached mean $F1_{sapro} = 0.97 \pm 0.01$ and $MCC = 0.96 \pm 0.02$ across the five folds on the *Salmo* data set,

while the corresponding *Oncorhynchus*-trained model achieved $F1_{sapro} = 0.80 \pm 0.09$ and MCC = 0.78 ± 0.10 on the *Oncorhynchus* data. Generalist EfficientNetV2-S models trained on "Taxa present in both classes" or "Taxa with ≥10 photographs in both classes" performed similarly on both datasets ($F1_{sapro} \approx 0.95–0.96$ on the *Salmo* data and ≈ 0.79 on the *Oncorhynchus* data, MCC ≈ 0.77–0.94). Because the genus-specific datasets were also used during model training, these scores involve substantial data re-use and should be interpreted as optimistic, exploratory upper bounds rather than unbiased test performance (Table S2).

3.2 Qualitative and qualitative analysis

Grad-CAM analysis revealed that surface reflections were consistently responsible for misclassifications for EfficientNetV2S, particularly when surface reflections obscured fish features (Figure 4). 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 (Fig. 4e). 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 (Fig. 4f). However, where water surface reflections created both bubbles above and reflective patterns below the fish (Figs 4c-d) the corresponding heatmap (Figure 4g-h) reveals that the model focused primarily on these water disturbances leading to misclassifying these healthy fish as diseased.

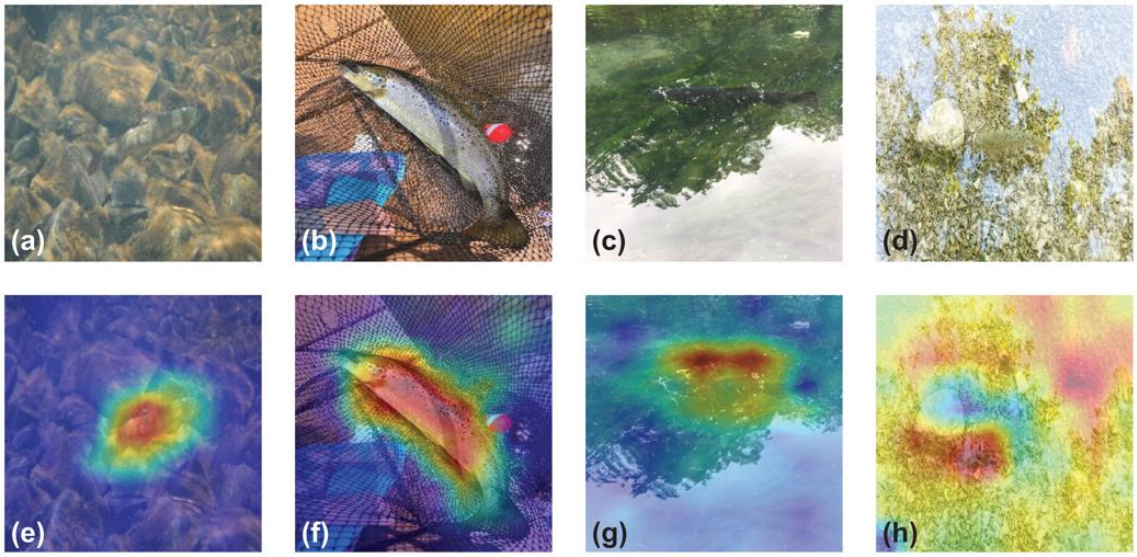


*Figure 4. Grad-CAM heatmaps. Comparison of a subset of images used for EfficientNetV2S classification on the 'Salmo, ≥10 photographs in both classes' dataset (a-d) and their corresponding Grad-CAM heatmap overlays (e-h). (a) Atlantic salmon (*Salmo salar*) correctly classified as infected with '*Saprolegnia *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 Section S2 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 Section S5.

4 Discussion

Our results demonstrate the potential for computer vision to support disease surveillance in wild fisheries. All models significantly outperformed random classification (Fig. 3), particularly 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, our manual inspection of the images used (*https://www.kaggle.com/datasets/subirbiswas19/freshwater-fish-disease-aquaculture-in-south-asia*), suggests 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.

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. Grad-CAM visualisations confirmed the model focused on relevant

anatomical features in correctly classified images. In cases of misclassification, however, the model's focus was often on water surface reflections rather than the fish itself (Fig. 4). This qualitative finding is supported by our quantitative analysis, which found no significant link between misclassifications and global image quality metrics like overall sharpness or perceptual quality (Section S5). It must be noted, however, that the statistical power of this quantitative analysis was inherently low due to the small number of misclassified images (14 'healthy', 3 '*Saprolegnia* spp.'). Despite this limitation, the combined evidence strongly suggests that model errors are not driven by generally 'poor quality' images, but by the specific, misleading artefacts like reflections identified by Grad-CAM. This provides clear guidance for improving image acquisition protocols or pre-processing steps in future work.

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 of computer vision as a surveillance tool is the resources required for classification. Model training required significant computational resources (148GB RAM and two Nvidia P100 GPUs), with processing time ranging from 15 to 47 hours depending on the model and dataset size. While EfficientNetV2S delivered the best performance, EfficientNetV2B0 offers a practical alternative with only one-third of the

parameters (Tan and Le 2021), enabling deployment on resource-constrained devices while maintaining strong performance. This trade-off between accuracy and efficiency is crucial for real-world implementation, particularly in field settings.

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 programs with standardised imaging protocols (August et al. 2020). 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 field-acquired images, providing a cost-effective alternative to traditional field surveys; they capture valuable metadata, including date, time, location, which is important for disease surveillance and in iNaturalist '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.

4.3 Challenges in Wild Fish Disease Detection

Beyond the challenge of specific image artefacts like surface reflections, the pipeline's reliance on image-level classification presents a key limitation. This approach prevents a per-fish assessment in images containing multiple individuals. Future work could address this limitation in two ways. A two-stage pipeline could use an object detector like 'megafishdetector' (Yang et al. 2023) to first locate and crop each fish, before our existing classifier is applied to each individual. A more comprehensive solution would be to develop a dedicated object detection model from the ground up, though this would require the significant effort of re-annotating the dataset with bounding boxes. Both approaches represent key directions for improving the granularity of this surveillance tool. Pre-processing techniques could reduce reflection effects crucial for minimally invasive *in situ* monitoring. It is clear from the results on the *Salmo* dataset that a narrower taxon focus can yield good results, so a hierarchical model incorporating taxonomic data (Elhamod et al. 2022), may improve classification accuracy.

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 *Saprolegenia* 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 exploratory analysis on the *Salmo* and *Oncorhynchus* genus-specific datasets (Table S2) further suggests that, although models trained on a specific genus tend to perform best on that genus, generalist models trained on broader salmonid data can approach this level of performance under conditions of substantial data re-use. This supports the idea that a single, broadly trained model may be sufficient for many surveillance applications, with targeted genus-specific fine-tuning reserved for high-priority host groups.

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 outbreaks for managers. Large-scale post-hoc analyses incorporating geotags and temporal data could also help identify consistent spatial and temporal patterns of disease occurrence, offering valuable ecological insights into outbreak dynamics. Second, as an *in situ* mobile application. Our model comparison was in part designed to explore the trade-offs for such on-device deployment. While an ultra-lightweight model like MobileNetV3S offers a rapid assessment that performs well above baseline (Figure 3) (around 27 ms per image in our tests), architectures like EfficientNetV2 are also suitable for mobile use and provide a significant increase in accuracy (Tan and Le 2021). EfficientNetV2B0 require roughly 100 ms per image and EfficientNetV2S about 400 ms (Table 4). This shows that a range of models are viable for a practical real-time, in-field tool, offering a choice between maximum efficiency and higher performance.

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 in terms of how to analyse the data and in interpreting any apparent patterns in disease occurrence or distribution.

5 Conclusion

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.

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].

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.

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.

August, T.A., Pescott, O.L., Joly, A. and Bonnet, P. 2020. 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.

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.

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.

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.

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.

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.

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.

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.

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.

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.

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.

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.

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.

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.

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

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.

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.

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.

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.

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.

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.

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.

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). Available at: https://www.mapress.com/zt/article/view/zootaxa.4896.2.6 [Accessed: 13 February 2024].

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.

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.

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.

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.

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.

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.

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.

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.

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.

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.

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.

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.

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.

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.

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.

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.

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.

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.

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].

Declaration of interests

☒ The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

☐ The authors declare the following financial interests/personal relationships which may be considered as potential competing interests:

Highlights

6. Computer vision pipeline detects *Saprolegnia* spp. disease signs in wild salmonids
7. Citizen science images viable source for fish disease detection
8. EfficientNetV2S achieves highest recall and precision for on *Salmo* spp. specific dataset
9. Model performance sensitive to host taxonomic level and image quality
10. AI enables large-scale non-invasive surveillance of wildlife disease