



HAL
open science

Machine learning is a powerful tool to study the effect of cancer on species and ecosystems

Antoine M Dujon, Marion Vittecoq, Georgina Bramwell, Frédéric Thomas,
Beata Ujvari

► To cite this version:

Antoine M Dujon, Marion Vittecoq, Georgina Bramwell, Frédéric Thomas, Beata Ujvari. Machine learning is a powerful tool to study the effect of cancer on species and ecosystems. *Methods in Ecology and Evolution*, 2021, 12 (12), pp.2310-2323. 10.1111/2041-210X.13703 . hal-04350061

HAL Id: hal-04350061

<https://hal.science/hal-04350061>

Submitted on 18 Dec 2023

HAL is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.



Distributed under a Creative Commons Attribution - NonCommercial 4.0 International License

Machine learning is a powerful tool to study the effect of cancer on species and ecosystems

Antoine M. Dujon^{1,2,3}  | Marion Vittecoq^{2,4,5}  | Georgina Bramwell^{1,3}  |
Frédéric Thomas^{2,3,4}  | Beata Ujvari^{1,3} 

¹Geelong, School of Life and Environmental Sciences, Centre for Integrative Ecology, Deakin University, Waurn Ponds, Victoria, Australia

²CREEC, UMR IRD 224-CNRS 5290-Université de Montpellier, Montpellier, France

³CANECEV-Centre de Recherches Ecologiques et Evolutives sur le cancer (CREEC), Montpellier, France

⁴MIVEGEC, University of Montpellier, CNRS, IRD, Montpellier, France

⁵Tour du Valat, Research Institute for the Conservation of Mediterranean Wetlands, Arles, France

Correspondence

Antoine M. Dujon
Email: antoine.dujon@yahoo.fr

Funding information

MAVA Foundation; ANR TRANSCAN

Handling Editor: Aaron Ellison

Abstract

1. Cancer is an understudied but important process in wildlife. Cancerous cells are proposed to have had significant effect on the evolution of metazoan species due to their negative effect on host fitness. However, gaining knowledge on the impact of cancer on species and ecosystems is currently relatively slow as it requires expertise in both ecology and oncology. The field can greatly benefit from automation to reduce the need of excessive manpower and analyse complex ecological datasets.
2. In this commentary, we examine how machine learning has been used to gain knowledge on oncogenic processes in wildlife. Using a landscape ecology approach, we explore spatial scales ranging from the size of a molecule up to whole ecosystems and detail, for each level, how machine learning has been used, or could contribute to obtain insights on cancer in wildlife populations and ecosystems.
3. We illustrate how machine learning is a powerful toolbox to conduct studies at the interface of ecology and oncology. We provide guidance for the readers of both fields on how to implement machine learning tools in their research and identify directions to move the field forward using this promising technology. We demonstrate how applying machine learning to complex ecological datasets will (a) contribute to quantitating the effect of cancer at different life stages in wildlife; (b) allow the mining of long-term datasets to understand the spatiotemporal variability of cancer risk factors and (c) contribute to mitigating cancer risk factors and the conservation of endangered species.
4. With this study, we aim to facilitate the use of machine learning to wildlife species and to encourage discussion between the scientists of the fields of oncology and ecology. We highlight the importance of international and pluridisciplinary collaborations to collect high-quality datasets on which efficient machine learning algorithms can be trained.

KEYWORDS

artificial intelligence, computer vision, epidemiology, remote sensing, species interactions, wildlife diseases

1 | INTRODUCTION

Cancer is a group of diseases found in multicellular organisms in which cheating cells break cellular cooperation, go rogue, and obtain the ability to evolve via Darwinian selection inside the host (Aktipis et al., 2015; Albuquerque et al., 2018). The emergence of cancerous cells is increasingly seen as a speciation event, in which a new parasitic species emerges, initiates a clade and consumes resources from the host, impairing the host's health and fitness (Capp & Thomas, 2020; Duesberg et al., 2011). In most cases, malignant cells die with their host (in contrast, see the nine cases of transmissible cancers; Dujon, Gatenby, et al., 2020; Ganguly et al., 2016; Loh et al., 2006; Metzger et al., 2016; Pye et al., 2016).

Species are exposed to cancer risk factors forming a dynamic landscape which can vary relatively fast in space and in time and can lead to an increase in cancer incidence (Dujon, Ujvari, et al., 2020; Pesavento et al., 2018). A typical example is the rapid change induced by human activities (Giraudeau et al., 2018; Pesavento et al., 2018), which increase exposure to carcinogenic substances in almost all ecosystems, intensifies habitat fragmentation and reduces genetic diversity, all of which escalates the likelihood to develop cancer in animals (see examples: sea turtles, Jones et al., 2016; belugas, Martineau et al., 2002; California sea lions, Browning et al., 2015).

The need for the collection of datasets to study oncogenic processes both in laboratories and in the wild is rapidly increasing (Dujon, Aktipis, et al., 2021). However, such studies are often laborious to conduct, involve the use of a range of different methods and technologies (e.g. genetics, histology, bio-logging), and often multiple experts trained in a range of disciplines (in oncology, behavioural ecology, chemistry; Dujon, Bramwell, et al., 2020; Schlemmer et al., 2018; Weinstein, 2018). In addition, many of those studies require tools able to make predictions efficiently (currently done mostly by human experts), with a real potential for facilitating them through automation. Indeed, the ability to learn the patterns present in complex ecological datasets will (a) contribute to the quantification of the effect of tumours present at various stages in wildlife species for which data are currently lacking; (b) allow for the reanalysis of long-term datasets (e.g. remote sensing, population censuses) to understand the spatiotemporal variability of cancer risk factors and predict future impacts; and (c) contribute to the mitigation of cancer risk factors to improve outcomes for the conservation of endangered species (Dujon, Ujvari, et al., 2020; Hamede et al., 2020; Vittecoq et al., 2013).

The increasing need for predictive modelling has driven recent advances in the use of machine learning algorithms in science (Dujon & Schofield, 2019; L'Heureux et al., 2017; Lecun et al., 2015). Machine learning performs predictive tasks by generalising information from data samples (Domingos, 2012; Esteva et al., 2021; L'Heureux et al., 2017; Lecun et al., 2015) and are used in situations in which robust statistical inference is often considered less important as having a robust predictive model (and often machine learning is used to annotate data which is subsequently analysed within the inferential framework; Bzdok et al., 2018). In the context of the

study of the ecology of cancer such prediction tasks include, for example, the detection of internal or external tumours in imagery data or individual cancerous cells in liquid biopsies or histology samples but also finding cancer in individuals based on metabolite profiles. In addition, machine learning can be used to determine which species an individual belongs to, annotate its behaviour, obtain information on its habitat, but also on its exposure to chemicals and other risk factors (Cardoso et al., 2018; Kourou et al., 2015; Kumar et al., 2015; Salem et al., 2017). Those tasks are often time-consuming and could be considerably hasten using machine learning tools to satisfy the increasing demand in oncological expertise.

Over the past decade, and with the emergence of deep learning, convolutional neural networks have gained a raising popularity to classify image samples (Esteva et al., 2021; Lecun et al., 2015). In contrast, support vector machines, random forests and K-means clustering algorithms are often used to solve a broader range of prediction tasks (i.e. using gene expression, metabolome or patient records data; Domingos, 2012; L'Heureux et al., 2017; Lecun et al., 2015), see Box 1. It is therefore time to assess how machine learning has the potential to facilitate studies investigating the impact of cancers on wildlife species and ecosystems. In addition, it is important to identify the challenges that these approaches need to overcome.

In this paper, we review how machine learning has been used so far to study cancer using examples from both human and wildlife species. We also provide future directions in which machine learning would greatly contribute to our improved quantification of the effect of cancer on ecosystems. Finally, we outline challenges and recommendations to the reader on how to implement machine learning algorithms to analyse data in their study system. With this publication, we aim to illustrate how machine learning can improve the feasibility of studies on cancer in wildlife species and ecosystems and highlight the suitability of machine learning for their management and conservation (Hamede et al., 2020). By demonstrating the applicability and benefit of machine learning techniques to study cancer in the wild, we also pave the path for future studies investigating other communicable and non-communicable wildlife diseases.

2 | APPLICATION OF MACHINE LEARNING FOR THE DETECTION OF CANCER IN WILDLIFE SPECIES

The landscape ecology of cancer is a dynamic and complex network of interactions between risk factors within nested spatial scales ranging from molecular mechanisms in a cell, through the individual, to the level of populations and entire ecosystems (Daoust et al., 2013; Dujon, Ujvari, et al., 2020). This network of interactions ultimately determines the emergence, progression or disappearance of tumours in a host. Here we detail how machine learning can be used to obtain information about the effect of cancer on wildlife species at four key spatial scales: the cell, the individual, the species and the ecosystem (Figure 1).

BOX 1 Popular machine learning methods in ecology and oncology

Unsupervised learning

The aim of an unsupervised machine learning algorithm is to automatically find patterns in a dataset (e.g. different categories or clusters, as opposed to a supervised algorithm which requires them; Domingos, 2012). A popular unsupervised machine learning algorithm is the K-means clustering algorithm which finds clusters of observations with similar characteristics in a dataset (Jain, 2010). Unsupervised machine learning algorithms are useful in exploratory analysis because they can identify patterns that may not be obvious in the raw data, as for example when analysing omic datasets (Valletta et al., 2017). In addition, unsupervised machine learning algorithms perform well when the groups present in the data are very distinct from each other (i.e. they can be useful in detecting anomalies such as tumour cells, or abnormal behaviour in tumour-bearing individuals). Recent advances in deep learning are currently making rapid progress towards fully unsupervised learning in the medical field which could be translated to wildlife species (Esteva et al., 2021).

Support vector machines

Support vector machines are supervised algorithms that construct a set of boundaries (i.e. hyperplanes) in an N-dimensional space (with N being the number of variables included in the model) to classify records in a dataset in multiple categories (Thessen, 2016). Support vector machines often perform relatively well when given labelled training data and have been used to predict susceptibility to cancer based on single nucleotide profiles (Kourou et al., 2015), classify histology samples (Komura & Ishikawa, 2018), predict habitat types (Mountrakis et al., 2011) and classify animal behaviours (Valletta et al., 2017).

Random forest

Random forests are supervised algorithms that operate by constructing a multitude of weak decision trees and by aggregating them to compute a prediction following a bootstrapping approach (hence the term random forest). The predictions returned by a random forest algorithm outperform those returned by each of the constituting trees. Random forests are one of the most widely used machine learning in ecology (Dujon & Schofield, 2019; Kampichler et al., 2010; Thessen, 2016; Valletta et al., 2017). They have also been used to classify cells as healthy or malignant based on their shape (Kumar et al., 2015), and map the habitat or the behaviour of species (Cutler et al., 2007; Valletta et al., 2017).

Artificial neural networks and convolutional neural networks

Artificial neural networks are supervised machine learning algorithms that mimic the way the human brain analyses and processes information and are mostly used to classify records in a dataset. In the last 10 years, the popularity of artificial neural networks has greatly increased due to their ability to efficiently solve classification tasks, analyse large amount of data and a wider availability in hardware able to perform their costly computation (Jordan & Mitchell, 2015; Lecun et al., 2015). Convolutional neural networks are a subclass of artificial neural networks dedicated to the processing and classification of imagery data (e.g. histology slides, animal pictures; Esteva et al., 2017; Minakshi et al., 2017; Wäldchen & Mäder, 2018; Weinstein, 2018). The most advanced convolutional neural network algorithms are currently able to classify the content of whole pictures and accurately delineate different regions within them (Gray et al., 2019; Redmon et al., 2015; Ren et al., 2015; Temitope Yekeen et al., 2020).

3 | AT THE CELLULAR LEVEL

3.1 | Analysis of omics datasets

Cancerous cells have an altered metabolism that allows them to survive as rogue cells in the host organism, often reactivating ancient 'genetic programmes' that are present in all metazoan cells (Capp & Thomas, 2020; Kaushik & DeBeradinis, 2018). This implies that biomarkers that distinguish cancerous cells from healthy ones can be isolated and used in diagnostics, as evidenced by studies in humans

(Kaushik & DeBeradinis, 2018). However, identifying suitable biomarkers to detect cancerous cells in liquid or solid biopsy samples of wildlife species is currently a key challenge that remains to be addressed (Dujon, Aktipis, et al., 2021; Hamede et al., 2020). Machine learning is increasingly used to study cancer-specific biomarkers, facilitating the processing of large datasets generated by the screening of the genome, epigenome and transcriptome and that could also be beneficial for wildlife species (Lee et al., 2020). For example, machine learning (random forests) was used to investigate devil facial tumour disease biomarkers in the serum of Tasmanian devils *Sarcophilus harrisii*, and

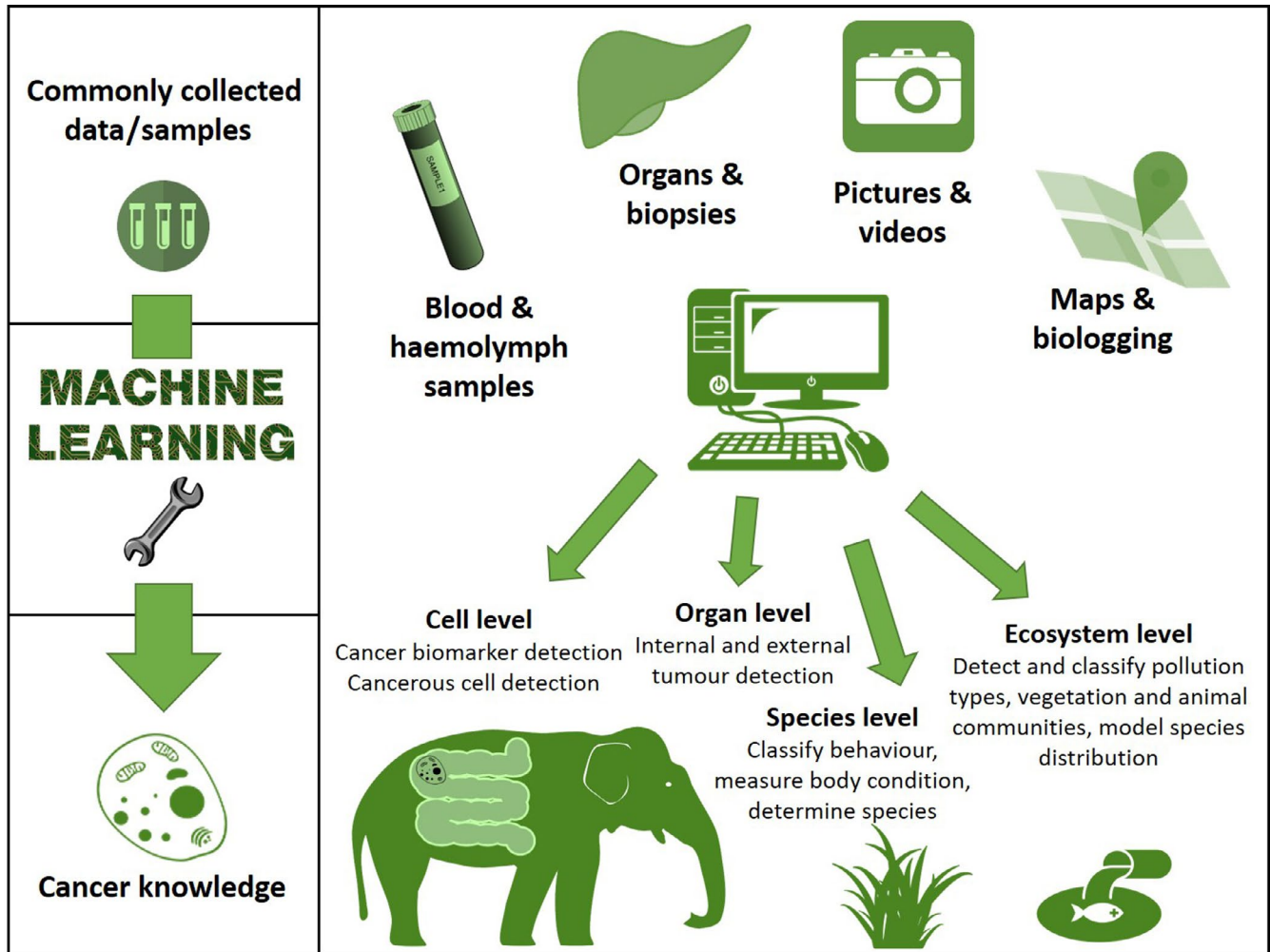


FIGURE 1 The potential of machine learning to obtain insights on the effect of cancer on species and ecosystems from the cellular to ecosystem scale

has facilitated the identification of a peptide biomarker specific to the transmissible cancer afflicting the species (Karu et al., 2016).

3.2 | Histology

The need for accuracy in histopathologic diagnosis of cancer is increasing as cancer therapy in humans requires accurate cancer type identification (Acs et al., 2020). Cancerous cells often have abnormal shapes with modified cell membranes or nuclei, and tend to cause disorganised cell arrangements which can be detected by a machine learning algorithm that is able to distinguish cancerous cells from healthy ones in histology samples (Komura & Ishikawa, 2018; Kumar et al., 2015, 2020). Within a species, machine learning algorithms are increasingly being used to automatically detect single or groups of malignant cells and determine if a sample contains cancerous cells (Schmidt et al., 2018; Sikpa et al., 2019). Such algorithms allow to annotate a large number of samples faster and often with comparable accuracy and precision compared to manual screening by a trained expert (Esteva et al., 2017; Hekler et al., 2019).

Designing machine learning algorithms that can detect cancerous cells in multiple species is challenging, but not impossible, and can be achieved using unsupervised anomaly detection models (e.g. using generative adversarial networks, Ren et al., 2019; Schlegl et al., 2019). Given enough samples, such models are likely to be able to identify different cell types which can then be verified by scientists to determine if they are neoplastic and iterated to improve the tuning of the model. This approach has been successfully applied to human samples and could be translated to wildlife species (Schlegl et al., 2017, 2019). Since the approach is unsupervised, it would allow the screening of species for which we have little information available as long as cancerous cells have a different observable phenotype compared to healthy ones.

3.3 | Detection of tumour cells in liquid biopsy

The detection of circulating tumour cells within the lymphatic system or blood vessels is currently challenging because their concentration is usually relatively low (e.g. in humans and dogs, Alix-Panabières &

Pantel, 2021; Crowley et al., 2013; Pantel & Alix-Panabières, 2010; Wright et al., 2019). Detecting circulating tumour cells in humans and mammals requires extremely sensitive and specific analytical methods. This classification can be done manually or using a machine learning algorithm to automate the process (Li et al., 2019; Wang et al., 2020). After an enrichment step used to concentrate cells into a small sample, each cell is classified as being healthy or cancerous using immunological, molecular or functional assays (Alix-Panabières & Pantel, 2021; Pantel & Alix-Panabières, 2010). Liquid biopsy provides promising opportunities for the analysis of blood samples from vertebrates (Chmielewska et al., 2013), or haemolymph from invertebrates, and will, for example, contribute to the detection of elusive transmissible cancers (Bramwell et al., 2021; Burioli et al., 2019). The technique is particularly useful for bivalves in which cancer is predicted to be a widespread and global disease (Dujon, Bramwell, et al., 2021), especially to detect enlarged, polyploid and abnormally shaped neoplastic cells in haemolymph samples. This would also facilitate long-term environmental monitoring and the acquisition of baseline data we are currently lacking (Kumar et al., 2015; Rodellar et al., 2018; Sikpa et al., 2019).

At the cellular level, machine learning could allow a time and cost-effective annotation of large sample collections accumulated by wildlife managers (especially blood) and veterinary practitioners (blood, histology and biopsy). For all methods described above, it is expected that the integration of machine learning into routine care will be a milestone for the healthcare sector in the coming decades, and subsequently will translate into improved ability to detect cancer in wildlife species (Acs et al., 2020).

4 | AT THE ORGAN LEVEL

4.1 | Internal and external solid tumours

At the organ level, the detection of solid tumours often relies on the identification of subtle differences between neoplastic and healthy tissues on imagery data (e.g. radiography, photography, computerised tomography scan). The analysis and interpretation of imaging data require trained subspecialists who are in high demand (Schlemmer et al., 2018). Recent advances in machine learning algorithms have made great strides in automatically detecting tumours in human medical imaging data (Bi et al., 2019; Esteva et al., 2021), with a performance that can even surpass human experts in task-specific applications (e.g. skin mole, lung and brain tumour detection, Akselrod-Ballin et al., 2016; Esteva et al., 2017).

The algorithms developed to detect cancer in humans can also be translated to the detection of tumours in animal imagery data, as illustrated by studies that used machine learning to detect cancer in the neck or tongues of mice (Lu et al., 2018; Ma et al., 2017) and classify skin tumours or to distinguish between meningiomas and gliomas in the brain of dogs (Banzato et al., 2018; Zapata et al., 2020). Imaging technologies coupled with machine learning also have the potential to allow precise volumetric delineation of

tumour size at multiple time points and to simultaneously track multiple lesions (Bi et al., 2019) and would also be particularly useful for detecting early-stage tumours in wildlife species. This data is very valuable to obtain insights on tumour progression, and to test novel hypotheses (e.g. the group phenotypic composition hypothesis; Capp et al., 2021) exploring the reasons why only a small fraction of tumours reach the metastatic stage (Dujon, Capp, et al., 2021; Luzzi et al., 1998). Indeed, theoretical modelling predicts that in the wild, animals with cancer will be rapidly eliminated because of the additional costs associated with the tumour burden (Perret et al., 2020; Vittecoq et al., 2013). This implies that it would be difficult to observe the late stages of metastatic cancers in most wild species (i.e. prey species) and it is of key importance to be able to detect precancerous lesions before an animal dies and ecologically relevant information is lost.

Precancerous lesions have already been observed in a number of species, for example, in the liver of flatfish or the digestive tube of sea turtles (Lima et al., 2019; Myers et al., 1991). In humans, precancerous lesions are often present in much larger numbers compared to cancerous ones (Bi et al., 2019; Folkman & Kalluri, 2004; Gould et al., 2015) and machine learning algorithms have been developed to distinguish between them to reduce over-diagnosis (Song et al., 2017). It is very likely that the same trends apply to wildlife species (Thomas et al., 2018). At the organism level, machine learning could take advantage of samples collected by non-specialists to enable managers, hunters and fishermen to perform routine monitoring of wildlife health.

5 | AT THE SPECIES LEVEL

5.1 | Species identification

The quantification of the effect of cancer risk factors in wildlife often requires the identification of the causative agent. This is, for example, the case for parasites (helminths, protozoans, arthropods and others) that are well-known causes of cancer in both human and wildlife populations (Brindley et al., 2015; Ewald & Swain Ewald, 2014; Jones et al., 2016). Machine learning can automate species identification, for example, by identifying helminth eggs in urine and stool samples (which could be applied to the stool of elusive or nocturnal species; Holmström et al., 2017), including cancer-causing helminth species such as *Schistosoma haematobium* (Brindley et al., 2015). Blood-sucking arthropods also contribute to cancer burdens in animal populations by carrying cancer-causing pathogens, and possibly even cancerous cells (see Arnal et al., 2020). Preliminary studies with promising results have investigated the potential of machine learning to create smartphone applications to identify mosquito species, with the potential to facilitate surveys in remote areas (Minakshi et al., 2017; Motta et al., 2019). Similarly, machine learning has been used to identify triatomine species (*Triatoma* sp, Khalighifar et al., 2019), the vectors of *Trypanosoma cruzi*, a parasite that has antiangiogenic and anti-tumoral properties (López et al., 2010).

In aquatic ecosystems, the frequency of phytoplankton blooms has been increasing, and the carcinogenic toxins produced by these events most likely contribute to elevated tumour burden in species inhabiting these ecosystems (Bláha et al., 2009; Hallegraeff, 1993). For example, sea turtles exposed to brevetoxins produced by red tide algal blooms have decreased body condition and increased oxidative stress levels that have been correlated with fibropapillomatosis tumour growth (Perrault et al., 2017). Embedded in portable flow cytometers, machine learning algorithms can identify toxic phytoplankton species in the field during algal blooms, and therefore, aid the monitoring of these toxic events (Göröcs et al., 2018). In the laboratory, machine learning can complement experts to identify oncogenic species in water samples (i.e. toxic dinoflagellate species; Culverhouse et al., 2003).

Quantify the effect of cancer on ecosystems will very likely require longitudinal studies in which multiple animals of a same species, with or without cancer, are being monitored over time (Dujon, Ujvari, et al., 2020; Ruiz-Aravena et al., 2018). Such studies require each individual to be repeatedly identified with a high accuracy which is currently primarily done manually. Promising studies (ranging from coral to mammals; Weinstein, 2018) indicate that machine learning can facilitate this task and could complement similar established methods to repeatedly identify individuals on commonly used devices such as camera traps, or video/sound recorders (Dujon, Ierodiaconou, et al., 2021; Siddiqui et al., 2018; Tabak et al., 2019; Wäldchen & Mäder, 2018; Weinstein, 2018).

5.2 | Behavioural classification

Alike parasitic infections, cancer is known to alter the behaviour of the host (Vittecoq et al., 2015). Humans with cancer often experience mood alteration, sleep disorders and a state of pain (Stepanski et al., 2009). Altered behaviour in response to cancer has also been observed in multiple animal species. For example, cancerous common fruit flies *Drosophila melanogaster* can discriminate between individuals at different stages of tumour development and selectively choose their social environment accordingly. Cancerous individuals kept in isolation exhibited faster tumour progression than flies kept in groups of other cancerous individuals (Dawson et al., 2018). In addition, female flies with tumours reproduced, on average, 2 days earlier than healthy ones to adapt their reproductive strategy in response to tumour progression (Arnal et al., 2017). Cancer in juvenile sea turtles *Chelonia mydas* increase the proportion of time they spend resting (Brill et al., 1995) and can trigger a basking behaviour not observed in healthy individuals (Swimmer, 2006). In Ukraine, great tits *Parus major* and flycatchers *Ficedula hypoleuca* avoid nest boxes located in areas heavily contaminated by radioactive elements (Møller & Mousseau, 2007), likely as a cancer avoidance strategy (Boutry et al., 2020).

Experiments investigating the effect of cancer on a host species usually involve the recording of behaviour by a human observer or with the help of cameras or other devices from which the

behavioural states can be inferred (Weinstein, 2018). However, the manual analysis of the data can be painstakingly time-consuming (Kain et al., 2013; Valletta et al., 2017; Weinstein, 2018). To speed up data analysis, machine learning models have been developed to automatically annotate the behaviour of a range of species such as hydras (Han et al., 2018), mice (de Chaumont et al., 2019) and drosophila (Kain et al., 2013) based on video records in laboratories. The use of such algorithms could be extended to study the effect of cancer on studied species and to create detailed behavioural maps that may allow the detection of subtle differences in the behaviour of healthy and tumour-bearing individuals (Berman et al., 2014).

In the wild, machine learning can be used to study the behaviour of animals and the space use of animal within the well-established field of movement ecology (Nathan et al., 2008). As species move in their cancer risk landscape, their exposure to carcinogenic substances will vary accordingly (Dujon, Ujvari, et al., 2020; Giraudeau et al., 2018; Pesavento et al., 2018). Tracking animals could potentially allow to measure their exposure to cancer risk factors but also detect and quantify alterations in their behaviour due to oncogenic processes. For example, a study on juvenile sea turtles *C. mydas* in Hawaii using acoustic telemetry found that individuals with fibropapillomatosis were less active during night time compared to healthy individuals, indicating that they spend more time resting on the seabed (possibly due to the energetic cost of tumour burden; Brill et al., 1995). Similarly, a tracking study of Tasmanian devils in Australia showed that the home range size of animals decreased since the apparition of a deadly transmissible cancer, likely because as the population size decreased, the per capita resources increased with animals requiring less space to acquire them (Comte et al., 2020). Here, machine learning can be used to considerably reduce the time required to annotate and predict the behaviour of animals tracked with a range of telemetry technology such as Global Positioning System (GPS) or accelerometers (Bidder et al., 2014; Chakravarty et al., 2019; Yu et al., 2021) and to obtain valuable information on species that can be otherwise difficult to monitor (e.g. Dujon et al., 2019; Schofield et al., 2017). For example, by providing both GPS location and behavioural states (using accelerometers) bio-logging will likely also be particularly useful to study the cancer risk landscape and infer the exposure to stressors of long-distance migratory species. In addition, machine learning models can often be easier to implement compared to more traditional animal movement models (Wijeyakulasuriya et al., 2020).

5.3 | Body condition

The energetic cost of bearing tumours increases as a cancer progresses and is often associated with a decrease in body condition in both humans and wild species (Friesen et al., 2015; Nicholson et al., 2018; Ruiz-Aravena et al., 2018). In addition to provide information on a species' life-history traits, the measurement of the body condition of individuals over time can be used as non-specific indication of the presence of a cancer (Nicholson et al., 2018) but also of

the progression of the tumour burden (see e.g. Ruiz-Aravena et al., 2018). However, it can be difficult to measure the body condition of elusive, endangered or simply overly heavy species. The use of emerging technologies, such as unmanned aerial vehicles, complemented by machine learning offers a promising avenue to measure the body condition of wildlife in a non-invasive way. This approach has already been applied to whales (Christiansen et al., 2016; Gray et al., 2019) and machine learning was used to automatically estimate the body condition of the animals in three species (Gray et al., 2019). This suggests the technique would be a valuable tool to obtain information for the beluga whales *Delphinapterus leucas* in which the prevalence of cancer is high (Martineau et al., 2002). Very similarly, the body condition of pinnipeds such as Australian sea lions *Neophoca cinerea*, and other pinniped species, can be inferred from photogrammetry (Allan et al., 2019; Hodgson et al., 2020), and machine learning automation could be implemented and translated to monitor species such as the California sea lion *Zalophus californianus* with a high prevalence of urogenital carcinoma (Browning et al., 2015).

6 | AT THE ECOSYSTEM LEVEL

The effect of cancer at the ecosystem level is currently understudied but can be significant (Vittecoq et al., 2013). For example, >80% of Tasmanian devils died over a short span of 25 years due to transmissible cancers. Devils being top predators and scavengers, their population decline has greatly affected the Tasmanian ecosystem (Cunningham et al., 2018; Hollings et al., 2014, 2016). Species inhabiting an ecosystem evolve in cancer risk landscapes. Such landscapes describe how spatiotemporal variability affect cumulative cancer risk exposure and the likelihood of species of developing cancer during their lifetime (Dujon, Ujvari, et al., 2020). Machine learning offers a range of algorithms commonly used to analyse remote sensing datasets facilitating the mapping of a range of environmental factors that influence the cancer risk landscape (Blaschke, 2010; Cutler et al., 2007; Dujon & Schofield, 2019; Kampichler et al., 2010).

6.1 | Habitat distribution

A number of cancer risk factors are differentially affected by the habitat type (exposure to ultraviolet, pollution, habitat fragmentation, see Dujon, Ujvari, et al. (2020) for a review). Machine learning has been used to map habitat types in a diverse range of ecosystems (Blaschke, 2010; Dujon & Schofield, 2019; Mountrakis et al., 2011; Qian et al., 2015) but also human-induced environmental degradation such as contamination by hydrocarbon spills (Temitope Yekeen et al., 2020), radioactive fallouts (Varley et al., 2016) as well as habitat fragmentation (Qian et al., 2015) or light pollution (Xue et al., 2020). The increased use of machine learning in habitat mapping will directly benefit the study of the effect of cancer on species and ecosystems.

6.2 | Species distribution

Determining how species interactions within an ecosystem (e.g. predator/prey interactions, grazing, parasitism, bioturbation, etc.) contribute to their cancer risk, and how the oncogenic processes associated with those taxa impact ecosystem functioning requires the determination of their spatial distribution. This is important as we are currently in a context of increasing oncogenic pressures on ecosystems and we have little knowledge of how biodiversity hotspots (or areas of high ecological values) overlap with area of high cancer risk (Dujon, Aktipis, et al., 2021; Vittecoq et al., 2013). Machine learning has been used to map the distribution of large number of plant and animal species (Dujon & Schofield, 2019; Tabak et al., 2019; Wäldchen & Mäder, 2018) including mosquitoes (Früh et al., 2018; Hardy et al., 2019) and helminths (Xia et al., 2019), the latter also being associated with increased cancer risk. In Hungary, Erdélyi et al. (2009) used a random forest algorithm to estimate the wider distribution area of an endemic papillomavirus infection associated with skin tumours in the Roe deer *Capreolus capreolus*. In Hawaii, Williams et al. (2010) used a very similar machine learning method to model the distribution of *Porites* growth anomalies, a tumour-like disease of coral. Machine learning has also aided in the identification of other potential animal reservoirs of zoonoses (Wardeh et al., 2020); the mapping of algal toxic blooms (Grasso et al., 2019); and has also been used to create metapopulation models and species interaction maps in the Tasmanian devil (Siska et al., 2018).

7 | MACHINE LEARNING IN APPLICATION

Current applications of machine learning clearly indicate the great potential of these techniques across different disciplines; however, a number of challenges need to be considered. First, we recommend strong collaborations between the field of ecology, oncology, and machine learning due to the fast development and expansion of the latter (Jordan & Mitchell, 2015). In addition, we also recommend that scientists outside the field of machine learning familiarise themselves with machine learning methods (in the same way they learn classical statistical analyses methods, see Olden et al., 2008; Domingos, 2012; Chicco, 2017; Dujon & Schofield, 2019 for an introduction to machine learning, see also Table 1). Before implementing a machine learning algorithm to study cancer, a scientist should consider several key points (Wagstaff, 2012). Machine learning, like any other statistical tool, is not an end by itself and can only generate new insights and add value to a study if it is associated with a good biological question to answer (Wagstaff, 2012). As such, machine learning would be especially useful if paired with one of the key questions recently identified in the field of ecology and evolution of cancer (see the list of 84 key questions from Dujon, Capp, et al., 2021), especially if it requires a large amount of automation (Figure 2). For example, such algorithms would facilitate the annotation of data which can then be included in inferential modelling approaches to improve our knowledge of how human activities

TABLE 1 Commonly used machine learning methods encountered in the field of ecology and evolution of cancer and suggested R and Python packages to implement them (adapted from Dujon & Schofield, 2019)

Machine learning algorithm	R packages	Python packages	Useful sources
Random forest	rf	scikit-learn	Cutler et al. (2007) and Strobl et al. (2009)
Support vector machine	e1071	scikit-learn	Hsu et al. (2008)
K-means clustering	Stats, ClusterR	scikit-learn	Jain (2010)
Convolutional and artificial neural network	keras	tensorflow, keras	Chollet (2015)
Imagery thresholding and segmentation	imager	OpenCV	Zou et al. (2016)

are modifying the prevalence of cancers in wildlife species (which is expected to increase as the number of contaminants released in ecosystems increases, Bramwell et al., 2021; Giraudeau et al., 2018; Pesavento et al., 2018), or to obtain data on the importance of cancer in ecosystem functioning (i.e. theoretical modelling predicts that prey with cancer are rapidly eliminated from the ecosystem by predators; Perret et al., 2020). Similarly machine learning algorithms can contribute to the detection of new transmissible cancer lineages that are elusive but predicted to be widespread (Capp et al., 2021; Dujon, Bramwell, et al., 2021; Metzger et al., 2016; Skazina et al., 2021).

Data availability is also an important consideration (Figure 2). Machine learning algorithms learn patterns from the data they are trained on; therefore, data quality largely determines model quality (Marx, 2019). Since cancer detection can be a relatively difficult task (especially in wildlife), training the algorithms to ease diagnostic could require relatively large datasets (Marx, 2019, Figure 2). The dataset size required to train a machine algorithm varies depending of the model type and of the task to solve (e.g. the number of samples required to train a convolutional neural network can be greatly reduced using transfer learning; Khan et al., 2019). An approach to determine the required size for a dataset is to split it and to train a machine learning algorithm by iteratively increasing the amount of data incorporated in the training process while monitoring how the increase in data increases your accuracy. At the cellular level, generation of high-quality reference datasets requires the compilation of genomic and proteomic information from a range of wildlife species, an ongoing task (see, e.g., the Ensembl repository which contains the genome of >300 vertebrate species, www.ensembl.org). At the organ level, histology datasets including samples from healthy and cancerous individuals must be compiled, for example using historical collections and then identifying needs for the collection of new data (e.g. Sikpa et al., 2019, trained their algorithm on 100 whole histology slides to detect tumours in mouse brains). In addition, it will also be important to share these datasets with the wider community to foster new ideas and refine approaches to train machine learning. Almost as important as creating reference datasets will be to share model architecture that performed well, but also that did underperformed, as to reduce the time required to develop efficient monitoring tools (Marx, 2019).

With this commentary, we have illustrated how machine learning is a powerful tool to add to the toolbox of both ecologists and oncologists. While the efforts required to train such machine learning algorithms will require initial large collaborations (as evidenced by

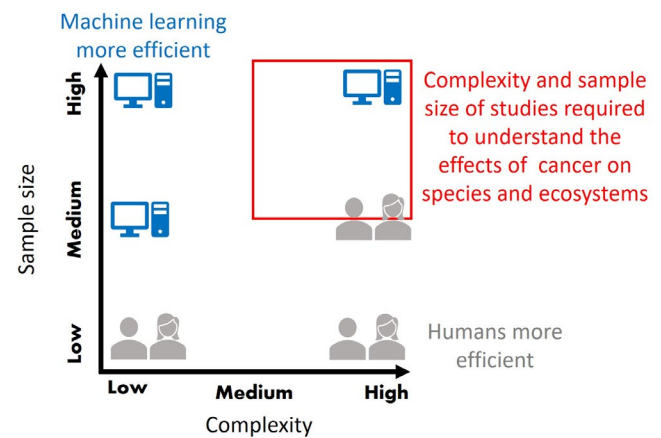


FIGURE 2 Diagram comparing the relative complexity and sample size of studies investigating the effects of cancer on species and ecosystems. A task can be considered complex because the data contains many dimensions (e.g. omics data), the properties of cancerous cells are close from those of healthy cells (e.g. certain type of tumours looks very similar on imagery data) or contain high spatiotemporal dependency (e.g. movement and other mapping data). The relative complexity and sample size of studies required to advance the field is represented as a red square. The human and computer icons represent areas of the diagram in which the analysis of the data by fully trained human specialists or by machine learning algorithm, respectively, would be in the majority of cases the most efficient solution. As the field progresses, it is expected that the complexity and sample size of studies will increase and will benefit from the advancements of machine learning. For convolutional neural networks, the sample size required to train a model can sometimes be considerably reduced using transfer learning

the recent advance in machine learning applied to human cancers), once properly developed they hold great potential for satisfying the increasing need for expert analyses. Answering these questions are of key importance to mitigate the effect of cancer on wildlife and implement efficient management strategies (Hamede et al., 2020).

ACKNOWLEDGEMENTS

This study was supported by the MAVA Foundation and the ANR TRANSCAN.

CONFLICT OF INTEREST

The authors have no conflict of interest to declare.

AUTHORS' CONTRIBUTIONS

A.M.D. wrote the original draft of the manuscript; A.M.D., M.V., G.B., B.U. and F.T. edited the manuscript; A.M.D. edited Figures 1 and 2. M.V. edited Figure 2. All authors contributed to the manuscript.

PEER REVIEW

The peer review history for this article is available at <https://publons.com/publon/10.1111/2041-210X.13703>.

DATA AVAILABILITY STATEMENT

No data were used in this publication.

ORCID

Antoine M. Dujon  <https://orcid.org/0000-0002-1579-9156>

Marion Vittecoq  <https://orcid.org/0000-0002-8930-6051>

Georgina Bramwell  <https://orcid.org/0000-0002-9883-4892>

Frédéric Thomas  <https://orcid.org/0000-0003-2238-1978>

Beata Ujvari  <https://orcid.org/0000-0003-2391-2988>

REFERENCES

- Acs, B., Rantalainen, M., & Hartman, J. (2020). Artificial intelligence as the next step towards precision pathology. *Journal of Internal Medicine*, 288(1), 62–81. <https://doi.org/10.1111/joim.13030>
- Akselrod-Ballin, A., Karlinsky, L., Alpert, S., Hasoul, S., Ben-Ari, R., & Barkan, E. (2016). A region based convolutional network for tumor detection and classification in breast mammography. In *Deep learning and data labeling for medical applications* (pp. 197–205). <https://doi.org/10.1007/978-3-319-46976-8>
- Aktipis, A. C., Boddy, A. M., Jansen, G., Hibner, U., Hochberg, M. E., Maley, C. C., & Wilkinson, G. S. (2015). Cancer across the tree of life: Cooperation and cheating in multicellularity. *Philosophical Transactions of the Royal Society B: Biological Sciences*, 370(1673), 20140219.
- Albuquerque, T. A. F., Drummond do Val, L., Doherty, A., & de Magalhães, J. P. (2018). From humans to hydra: Patterns of cancer across the tree of life. *Biological Reviews*, 93(3), 1715–1734. <https://doi.org/10.1111/brv.12415>
- Alix-Panabières, C., & Pantel, K. (2021). Liquid biopsy: From discovery to clinical application. *Cancer Discovery*, 11(4), 858–873. <https://doi.org/10.1158/2159-8290.CD-20-1311>
- Allan, B. M., Ierodionou, D., Hoskins, A. J., & Arnould, J. P. Y. (2019). A rapid UAV method for assessing body condition in fur seals. *Drones*, 3(1), 24. <https://doi.org/10.3390/drones3010024>
- Arnal, A., Jacqueline, C., Ujvari, B., Leger, L., Moreno, C., Faugere, D., Boidin-Wichlacz, C., Misse, D., Renaud, F., Montagne, J., Casali, A., Roche, B., Mery, F., & Thomas, F. (2017). Cancer brings forward oviposition in the fly *Drosophila melanogaster*. *Ecology and Evolution*, 7(1), 272–276. <https://doi.org/10.1002/ece3.2571>
- Arnal, A., Roche, B., Gouagna, L.-C., Dujon, A., Ujvari, B., Corbel, V., Remoue, F., Poinson, A., Pompon, J., Giraudeau, M., Simard, F., Missé, D., Lefèvre, T., & Thomas, F. (2020). Cancer and mosquitoes – An unsuspected close connection. *Science of the Total Environment*, 140631. <https://doi.org/10.1016/j.scitotenv.2020.140631>
- Banzato, T., Bernardini, M., Cherubini, G. B., & Zotti, A. (2018). A methodological approach for deep learning to distinguish between meningiomas and gliomas on canine MR-images. *BMC Veterinary Research*, 14(1), 1–6. <https://doi.org/10.1186/s12917-018-1638-2>
- Berman, G. J., Choi, D. M., Bialek, W., & Shaevitz, J. W. (2014). Mapping the stereotyped behaviour of freely moving fruit flies. *Journal of the Royal Society Interface*, 11(99), 20140672. <https://doi.org/10.1098/rsif.2014.0672>
- Bi, W. L., Hosny, A., Schabath, M. B., Giger, M. L., Birkbak, N. J., Mehrtash, A., Allison, T., Arnaout, O., Abbosh, C., Dunn, I. F., Mak, R. H., Tamimi, R. M., Tempany, C. M., Swanton, C., Hoffmann, U., Schwartz, L. H., Gillies, R. J., Huang, R. Y., & Aerts, H. J. W. L. (2019). Artificial intelligence in cancer imaging: Clinical challenges and applications. *CA: A Cancer Journal for Clinicians*, 1–31. <https://doi.org/10.3322/caac.21552>
- Bidder, O. R., Campbell, H. A., Gómez-Laich, A., Urgé, P., Walker, J., Cai, Y., Gao, L., Quintana, F., & Wilson, R. P. (2014). Love thy neighbour: Automatic animal behavioural classification of acceleration data using the k-nearest neighbour algorithm. *PLoS ONE*, 9(2), e88609. <https://doi.org/10.1371/journal.pone.0088609>
- Bláha, L., Babica, P., & Maršálek, B. (2009). Toxins produced in cyanobacterial water blooms - toxicity and risks. *Interdisciplinary Toxicology*, 2(2), 36–41. <https://doi.org/10.2478/v10102-009-0006-2>
- Blaschke, T. (2010). Object based image analysis for remote sensing. *ISPRS Journal of Photogrammetry and Remote Sensing*, 65(1), 2–16. <https://doi.org/10.1016/j.isprsjprs.2009.06.004>
- Boutry, J., Dujon, A. M., Gerard, A.-L., Tissot, S., Macdonald, N., Schultz, A., Biro, P. A., Beckmann, C., Hamede, R., Hamilton, D. G., Giraudeau, M., Ujvari, B., & Thomas, F. (2020). Ecological and evolutionary consequences of anticancer adaptations. *iScience*, 23(11), 101716. <https://doi.org/10.1016/j.isci.2020.101716>
- Bramwell, G., Schultz, A. G., Sherman, C. D. H., Giraudeau, M., Thomas, F., Ujvari, B., & Dujon, A. M. (2021). A review of the potential effects of climate change on disseminated neoplasia with an emphasis on efficient detection in marine bivalve populations. *Science of the Total Environment*, 775, 145134. <https://doi.org/10.1016/j.scitotenv.2021.145134>
- Brill, R. W., Balazs, G. H., Holland, K. N., Chang, R. K., Sullivan, S., & George, J. C. (1995). Daily movements, habitat use, and submergence intervals of normal and tumor-bearing juvenile green turtles (*Chelonia mydas* L.) within a foraging area in the Hawaiian Islands. *Journal of Experimental Marine Biology and Ecology*, 185(2), 203–218. [https://doi.org/10.1016/0022-0981\(94\)00146-5](https://doi.org/10.1016/0022-0981(94)00146-5)
- Brindley, P. J., Da Costa, J. M. C., & Sripa, B. (2015). Why does infection with some helminths cause cancer? *Trends in Cancer*, 1(3), 174–182. <https://doi.org/10.1016/j.trecan.2015.08.011>
- Browning, H. M., Gulland, F. M. D., Hammond, J. A., Colegrove, K. M., & Hall, A. J. (2015). Common cancer in a wild animal: The California sea lion (*Zalophus californianus*) as an emerging model for carcinogenesis. *Philosophical Transactions of the Royal Society B: Biological Sciences*, 370(1673), 20140228. <https://doi.org/10.1098/rstb.2014.0228>
- Burioli, E., Trancart, S., Simon, A., Bernard, I., Charles, M., Oden, E., Bierne, N., & Houssin, M. (2019). Implementation of various approaches to study the prevalence, incidence and progression of disseminated neoplasia in mussel stocks. *Journal of Invertebrate Pathology*, 168(April), 107271. <https://doi.org/10.1016/j.jip.2019.107271>
- Bzdok, D., Altman, N., & Krzywinski, M. (2018). Statistics versus machine learning. *Nature Methods*, 15(4), 233–234. <https://doi.org/10.1038/nmeth.4642>
- Capp, J.-P., DeGregori, J., Nedelcu, A. M., Dujon, A. M., Boutry, J., Pujol, P., Alix-Panabières, C., Hamede, R., Roche, B., Ujvari, B., Marusyk, A., Gatenby, R., & Thomas, F. (2021). Group phenotypic composition in cancer. *eLife*, 10, 1–20. <https://doi.org/10.7554/eLife.63518>
- Capp, J. P., & Thomas, F. (2020). A similar speciation process relying on cellular stochasticity in microbial and cancer cell populations. *iScience*, 23(9), 101531. <https://doi.org/10.1016/j.isci.2020.101531>
- Cardoso, S., Baptista, D., Santos, R., & Rocha, M. (2018). A review on metabolomics data analysis for cancer applications. In *Practical Applications of Computational Biology and Bioinformatics, 12th International Conference* (pp. 157–165). Springer International Publishing. <https://doi.org/10.1007/978-3-319-98702-6>
- Chakravarty, P., Cozzi, G., Ozgul, A., & Aminian, K. (2019). A novel biomechanical approach for animal behaviour recognition using

- accelerometers. *Methods in Ecology and Evolution*, 10(6), 802–814. <https://doi.org/10.1111/2041-210X.13172>
- Chicco, D. (2017). Ten quick tips for machine learning in computational biology. *BioData Mining*, 10(1), 1–17. <https://doi.org/10.1186/s13040-017-0155-3>
- Chmielewska, M., Łosiewicz, K., Socha, P., Mecik-Kronenberg, T., & Wasowicz, K. (2013). The application of circulating tumor cells detecting methods in veterinary oncology. *Polish Journal of Veterinary Sciences*, 16(1), 141–151. <https://doi.org/10.2478/pjvs-2013-0022>
- Chollet, F. (2015). *Keras*. GitHub. Retrieved from <https://github.com/fchollet/keras>
- Christiansen, F., Dujon, A. M., Sprogis, K. R., Arnould, J. P. Y., & Bejder, L. (2016). Noninvasive unmanned aerial vehicle provides estimates of the energetic cost of reproduction in humpback whales. *Ecosphere*, 7(10), e01468. <https://doi.org/10.1002/ecs2.1468>
- Comte, S., Carver, S., Hamede, R., & Jones, M. (2020). Changes in spatial organization following an acute epizootic: Tasmanian devils and their transmissible cancer. *Global Ecology and Conservation*, 22, e00993. <https://doi.org/10.1016/j.gecco.2020.e00993>
- Crowley, E., Di Nicolantonio, F., Loupakis, F., & Bardelli, A. (2013). Liquid biopsy: Monitoring cancer-genetics in the blood. *Nature Reviews Clinical Oncology*, 10(8), 472–484. <https://doi.org/10.1038/nrclinonc.2013.110>
- Culverhouse, P. F., Williams, R., Reguera, B., Herry, V., & Gonzalez Gil, S. (2003). Do experts make mistakes? A comparison of human and machine labelling of dinoflagellates. *Marine Ecology Progress Series*, 247, 17–25.
- Cunningham, C. X., Johnson, C. N., Barmuta, L. A., Hollings, T., Woehler, E. J., & Jones, M. E. (2018). Top carnivore decline has cascading effects on scavengers and carrion persistence. *Proceedings of the Royal Society B: Biological Sciences*, 285, 20181582.
- Cutler, D. R., Edwards, T. C., Beard, K. H., Cutler, A., Hess, K. T., Gibson, J., & Lawler, J. J. (2007). Random forests for classification in ecology. *Ecology*, 88(11), 2783–2792. <https://doi.org/10.1890/07-0539.1>
- Daoust, S. P., Fahrig, L., Martin, A. E., & Thomas, F. (2013). From forest and agro-ecosystems to the microecosystems of the human body: What can landscape ecology tell us about tumor growth, metastasis, and treatment options? *Evolutionary Applications*, 6(1), 82–91. <https://doi.org/10.1111/eva.12031>
- Dawson, E. H., Bailly, T. P. M., Dos Santos, J., Moreno, C., Devilliers, M., Maroni, B., Sueur, C., Casali, A., Ujvari, B., Thomas, F., Montagne, J., & Mery, F. (2018). Social environment mediates cancer progression in *Drosophila*. *Nature Communications*, 9(1). <https://doi.org/10.1038/s41467-018-05737-w>
- de Chaumont, F., Ey, E., Torquet, N., Lagache, T., Dallongeville, S., Imbert, A., Legou, T., Le Sourd, A.-M., Faure, P., Bourgeron, T., & Olivio-Marin, J.-C. (2019). Real-time analysis of the behaviour of groups of mice via a depth-sensing camera and machine learning. *Nature Biomedical Engineering*, 3(11), 930–942. <https://doi.org/10.1038/s41551-019-0396-1>
- Domingos, P. (2012). A few useful things to know about machine learning. *Communications of the ACM*, 55(10), 78. <https://doi.org/10.1145/2347736.2347755>
- Duesberg, P., Mandrioli, D., McCormack, A., & Nicholson, J. M. (2011). Is carcinogenesis a form of speciation? *Cell Cycle*, 10(13), 2100–2114. <https://doi.org/10.4161/cc.10.13.16352>
- Dujon, A. M., Aktipis, A., Alix-Panabières, C., Amend, S. R., Boddy, A. M., Brown, J. S., Capp, J.-P., DeGregori, J., Ewald, P., Gatenby, R., Gerlinger, M., Giraudeau, M., Hamede, R. K., Hansen, E., Kareva, I., Maley, C. C., Marusyk, A., McGranahan, N., Metzger, M. J., ... Ujvari, B. (2021). Identifying key questions in the ecology and evolution of cancer. *Evolutionary Applications*, 14(4), 877–892. <https://doi.org/10.1111/eva.13190>
- Dujon, A. M., Bramwell, G., Raven, N., Schofield, G., Hamede, R., Thomas, F., & Ujvari, B. (2020). A global meta-analysis of over 50 years of multidisciplinary and international collaborations on transmissible cancers. *Evolutionary Applications*, 13(7), 1745–1755.
- Dujon, A. M., Bramwell, G., Roche, B., Thomas, F., & Ujvari, B. (2021). Transmissible cancers in mammals and bivalves: How many examples are there? *BioEssays*, 43(3), 2000222. <https://doi.org/10.1002/bies.202000222>
- Dujon, A. M., Capp, J.-P., Brown, J. S., Pujol, P., Gatenby, R. A., Ujvari, B., Alix-Panabières, C., & Thomas, F. (2021). Is There one key step in the metastatic cascade? *Cancers*, 13(15), 3693. <https://doi.org/10.3390/cancers13153693>
- Dujon, A. M., Gatenby, R. A., Bramwell, G., MacDonald, N., Dohrmann, E., Raven, N., Schultz, A., Hamede, R., Gérard, A.-L., Giraudeau, M., Thomas, F., & Ujvari, B. (2020). Transmissible cancers in an evolutionary perspective. *iScience*, 23(7), 101269. <https://doi.org/10.1016/j.isci.2020.101269>
- Dujon, A. M., Ierodiaconou, D., Geeson, J. J., Arnould, J. P. Y., Allan, B. M., Katsedilis, K. A., & Schofield, G. (2021). Machine learning to detect marine animals on UAV imagery: Effect of morphology, spacing, behaviour and habitat. *Remote Sensing in Ecology and Evolution*, <https://doi.org/10.1002/rse2.205>
- Dujon, A. M., & Schofield, G. (2019). Importance of machine learning for enhancing ecological studies using information-rich imagery. *Endangered Species Research*, 39, 91–104. <https://doi.org/10.3354/esr00958>
- Dujon, A. M., Stieglitz, T. C., Amice, E., & Webber, D. M. (2019). Snail leaps and bounds: Drivers of the diel movement pattern of a large invertebrate, the Caribbean queen conch (*Lobatus gigas*), in a marginal inshore habitat. *Canadian Journal of Zoology*, 97, 436–445.
- Dujon, A. M., Ujvari, B., & Thomas, F. (2020). Cancer risk landscapes: A framework to study cancer in ecosystems. *Science of the Total Environment*, 142955. <https://doi.org/10.1016/j.scitotenv.2020.142955>
- Erdélyi, K., Dencso, L., Lehoczki, R., Heltai, M., Sonkoly, K., Csányi, S., & Solymosi, N. (2009). Endemic papillomavirus infection of roe deer (*Capreolus capreolus*). *Veterinary Microbiology*, 138(1–2), 20–26. <https://doi.org/10.1016/j.vetmic.2009.02.002>
- Esteva, A., Chou, K., Yeung, S., Naik, N., Madani, A., Mottaghi, A., Liu, Y., Topol, E., Dean, J., & Socher, R. (2021). Deep learning-enabled medical computer vision. *Npj Digital Medicine*, 4(1), 5. <https://doi.org/10.1038/s41746-020-00376-2>
- Esteva, A., Kuprel, B., Novoa, R. A., Ko, J., Swetter, S. M., Blau, H. M., & Thrun, S. (2017). Dermatologist-level classification of skin cancer with deep neural networks. *Nature*, 542(7639), 115–118. <https://doi.org/10.1038/nature21056>
- Ewald, P. W., & Swain Ewald, H. A. (2014). Joint infectious causation of human cancers. *Advances in Parasitology*, 84, 1–26. <https://doi.org/10.1016/B978-0-12-800099-1.00001-6>
- Folkman, J., & Kalluri, R. (2004). Cancer without disease. *Nature*, 427, 787. <https://doi.org/10.1038/427787a>
- Friesen, D. E., Baracos, V. E., & Tuszynski, J. A. (2015). Modeling the energetic cost of cancer as a result of altered energy metabolism: Implications for cachexia. *Theoretical Biology and Medical Modelling*, 12(1), 1–18. <https://doi.org/10.1186/s12976-015-0015-0>
- Fruh, L., Kampen, H., Kerkow, A., Schaub, G. A., Walther, D., & Wieland, R. (2018). Modelling the potential distribution of an invasive mosquito species: Comparative evaluation of four machine learning methods and their combinations. *Ecological Modelling*, 388(November 2017), 136–144. <https://doi.org/10.1016/j.ecolmodel.2018.08.011>
- Ganguly, B., Das, U., & Das, A. K. (2016). Canine transmissible venereal tumour: A review. *Veterinary and Comparative Oncology*, 14(1), 1–12. <https://doi.org/10.1111/vco.12060>
- Giraudeau, M., Sepp, T., Ujvari, B., Ewald, P. W., & Thomas, F. (2018). Human activities might influence oncogenic processes in wild animal populations. *Nature Ecology and Evolution*, 2(7), 1065–1070. <https://doi.org/10.1038/s41559-018-0558-7>

- Gould, M. K., Tang, T., Liu, I.-L., Lee, J., Zheng, C., Danforth, K. N., Kosco, A. E., Di Fiore, J. L., & Suh, D. E. (2015). Recent trends in the identification of incidental pulmonary nodules. *American Journal of Respiratory and Critical Care Medicine*, 192(10), 1208–1214. <https://doi.org/10.1164/rccm.201505-0990OC>
- Grasso, I., Archer, S. D., Burnell, C., Tupper, B., Rauschenberg, C., Kanwit, K., & Record, N. R. (2019). The hunt for red tides: Deep learning algorithm forecasts shellfish toxicity at site scales in coastal Maine. *Ecosphere*, 10(12), <https://doi.org/10.1002/ecs2.2960>
- Gray, P. C., Bierlich, K. C., Mantell, S. A., Friedlaender, A. S., Goldbogen, J. A., & Johnston, D. W. (2019). Drones and convolutional neural networks facilitate automated and accurate cetacean species identification and photogrammetry. *Methods in Ecology and Evolution*, 10(9), 1490–1500. <https://doi.org/10.1111/2041-210X.13246>
- Göröcs, Z., Tamamitsu, M., Bianco, V., Wolf, P., Roy, S., Shindo, K., Yanny, K., Wu, Y., Koydemir, H. C., Rivenson, Y., & Ozcan, A. (2018). A deep learning-enabled portable imaging flow cytometer for cost-effective, high-throughput, and label-free analysis of natural water samples. *Light: Science and Applications*, 7(1). <https://doi.org/10.1038/s41377-018-0067-0>
- Hallegraef, G. M. (1993). A review of harmful algal blooms and their apparent global increase. *Phycologia*, 32(2), 79–99. <https://doi.org/10.2216/i0031-8884-32-2-79.1>
- Hamede, R., Owen, R., Siddle, H., Peck, S., Jones, M., Dujon, A. M., Giraudeau, M., Roche, B., Ujvari, B., & Thomas, F. (2020). The ecology and evolution of wildlife cancers: Applications for management and conservation. *Evolutionary Applications*, 13(7), 1719–1732. <https://doi.org/10.1111/eva.12948>
- Han, S., Taralova, E., Dupre, C., & Yuste, R. (2018). Comprehensive machine learning analysis of Hydra behavior reveals a stable basal behavioral repertoire. *eLife*, 7, 1–26. <https://doi.org/10.7554/eLife.32605>
- Hardy, A., Ettritch, G., Cross, D., Bunting, P., Liywali, F., Sakala, J., Silumesii, A., Singini, D., Smith, M., Willis, T., & Thomas, C. (2019). Automatic detection of open and vegetated water bodies using Sentinel 1 to map African malaria vector mosquito breeding habitats. *Remote Sensing*, 11(5). <https://doi.org/10.3390/rs11050593>
- Hekler, A., Utikal, J. S., Enk, A. H., Solass, W., Schmitt, M., Klode, J., Schadendorf, D., Sondermann, W., Franklin, C., Bestvater, F., Flaig, M. J., Krahl, D., von Kalle, C., Fröhling, S., & Brinker, T. J. (2019). Deep learning outperformed 11 pathologists in the classification of histopathological melanoma images. *European Journal of Cancer*, 118, 91–96. <https://doi.org/10.1016/j.ejca.2019.06.012>
- Hodgson, J. C., Holman, D., Terauds, A., Koh, L. P., & Goldsworthy, S. D. (2020). Rapid condition monitoring of an endangered marine vertebrate using precise, non-invasive morphometrics. *Biological Conservation*, 242, 108402. <https://doi.org/10.1016/j.biocon.2019.108402>
- Hollings, T., Jones, M., Mooney, N., & McCallum, H. (2014). Trophic cascades following the disease-induced decline of an apex predator, the Tasmanian devil. *Conservation Biology*, 28(1), 63–75.
- Hollings, T., Jones, M., Mooney, N., & McCallum, H. (2016). Disease-induced decline of an apex predator drives invasive dominated states and threatens biodiversity. *Ecology*, 97(2), 394–405. <https://doi.org/10.1890/15-0204.1>
- Holmström, O., Linder, N., Ngasala, B., Mårtensson, A., Linder, E., Lundin, M., Möllanen, H., Suutala, A., & Diwan, V., & Lundin, J. (2017). Point-of-care mobile digital microscopy and deep learning for the detection of soil-transmitted helminths and *Schistosoma haematobium*. *Global Health Action*, 10(3), <https://doi.org/10.1080/16549716.2017.1337325>
- Hsu, C.-W., Chang, C.-C., & Chih-Jen, L. (2008). A practical guide to support vector classification. *BJU International*, 101(1), 1396–1400.
- Jain, A. K. (2010). Data clustering: 50 years beyond K-means. *Pattern Recognition Letters*, 31(8), 651–666. <https://doi.org/10.1016/j.patrec.2009.09.011>
- Jones, K., Ariel, E., Burgess, G., & Read, M. (2016). A review of fibropapillomatosis in Green turtles (*Chelonia mydas*). *Veterinary Journal*, 212, 48–57. <https://doi.org/10.1016/j.tvjl.2015.10.041>
- Jordan, M. I., & Mitchell, T. M. (2015). Machine learning: Trends, perspectives, and prospects. *Science*, 349(6245), 255–260.
- Kain, J., Stokes, C., Gaudry, Q., Song, X., Foley, J., Wilson, R., & De Bivort, B. (2013). Leg-tracking and automated behavioural classification in *Drosophila*. *Nature Communications*, 4(1910). <https://doi.org/10.1038/ncomms2908>
- Kampichler, C., Wieland, R., Calmé, S., Weissenberger, H., & Arriaga-Weiss, S. (2010). Classification in conservation biology: A comparison of five machine-learning methods. *Ecological Informatics*, 5(6), 441–450. <https://doi.org/10.1016/j.ecoinf.2010.06.003>
- Karu, N., Wilson, R., Hamede, R., Jones, M., Woods, G. M., Hilder, E. F., & Shellie, R. A. (2016). Discovery of biomarkers for tasmanian devil cancer (DFTD) by metabolic profiling of serum. *Journal of Proteome Research*, 15(10), 3827–3840. <https://doi.org/10.1021/acs.jproteome.6b00629>
- Kaushik, A. K., & DeBeradinis, R. J. (2018). Applications of metabolomics to study cancer metabolism Akash. *Biochimica et Biophysica Acta-Reviews on Cancer*, 19870(1), 2–14. <https://doi.org/10.1016/j.bbcan.2018.04.009>. Applications
- Khalighifar, A., Komp, E., Ramsey, J. M., Gurgel-Gonçalves, R., & Peterson, A. T. (2019). Deep learning algorithms improve automated identification of Chagas disease vectors. *Journal of Medical Entomology*, 56(5), 1404–1410. <https://doi.org/10.1093/jme/tjz065>
- Khan, S. U., Islam, N., Jan, Z., Ud Din, I., & Rodrigues, J. J. P. C. (2019). A novel deep learning based framework for the detection and classification of breast cancer using transfer learning. *Pattern Recognition Letters*, 125, 1–6. <https://doi.org/10.1016/j.patrec.2019.03.022>
- Komura, D., & Ishikawa, S. (2018). Machine learning methods for histopathological image analysis. *Computational and Structural Biotechnology Journal*, 16, 34–42. <https://doi.org/10.1016/j.csbj.2018.01.001>
- Kourou, K., Exarchos, T. P., Exarchos, K. P., Karamouzis, M. V., & Fotiadis, D. I. (2015). Machine learning applications in cancer prognosis and prediction. *Computational and Structural Biotechnology Journal*, 13, 8–17. <https://doi.org/10.1016/j.csbj.2014.11.005>
- Kumar, A., Singh, S. K., Saxena, S., Lakshmanan, K., Sangaiah, A. K., Chauhan, H., Shrivastava, S., & Singh, R. K. (2020). Deep feature learning for histopathological image classification of canine mammary tumors and human breast cancer. *Information Sciences*, 508, 405–421. <https://doi.org/10.1016/j.ins.2019.08.072>
- Kumar, R., Srivastava, R., & Srivastava, S. (2015). Detection and classification of cancer from microscopic biopsy images using clinically significant and biologically interpretable features. *Journal of Medical Engineering*, 2015, 1–14. <https://doi.org/10.1155/2015/457906>
- L'Heureux, A., Grolinger, K., Elyamany, H. F., & Capretz, M. A. M. (2017). Machine learning with big data: Challenges and approaches. *Institute of Electrical and Electronics Access*, 5, 7776–7797. <https://doi.org/10.1109/ACCESS.2017.2696365>
- Lecun, Y., Bengio, Y., & Hinton, G. (2015). Deep learning. *Nature*, 521(7553), 436–444.
- Lee, D., Park, Y., & Kim, S. (2020). Towards multi-omics characterization of tumor heterogeneity: A comprehensive review of statistical and machine learning approaches. *Briefings in Bioinformatics*, 00(June), 1–19. <https://doi.org/10.1093/bib/bbaa188>
- Li, Y., Mahjoubfar, A., Chen, C. L., Niazi, K. R., Pei, L., & Jalali, B. (2019). Deep cytometry: Deep learning with real-time inference in cell sorting and flow cytometry. *Scientific Reports*, 9(1), 11088. <https://doi.org/10.1038/s41598-019-47193-6>

- Lima, S. R., Barbosa, J. M. D. S., Saracchini, P. G. V., Padilha, F. G. F., Leite, J. D. S., & Ferreira, A. M. R. (2019). Gastric lesions in free-living sea turtles: An underestimated disease that reflects the health of the ecosystem. *Science of the Total Environment*, 697, 133970. <https://doi.org/10.1016/j.scitotenv.2019.133970>
- Loh, R., Bergfeld, J., Hayes, D., O'Hara, A., Pyecroft, S., Raidal, S., & Sharpe, R. (2006). The pathology of devil facial tumor disease (DFTD) in Tasmanian Devils (*Sarcophilus harrisii*). *Veterinary Pathology*, 43(6), 890–895.
- López, N. C., Valck, C., Ramírez, G., Rodríguez, M., Ribeiro, C., Orellana, J., Maldonado, I., Albin, A., Anaconda, D., Lemus, D., Aguilar, L., Schwaeble, W., & Ferreira, A. (2010). Antiangiogenic and antitumor effects of *Trypanosoma cruzi* Calreticulin. *PLoS Neglected Tropical Diseases*, 4(7), 1–9. <https://doi.org/10.1371/journal.pntd.0000730>
- Lu, G., Wang, D., Qin, X., Muller, S., Wang, X. U., Chen, A. Y., Chen, Z. G., & Fei, B. (2018). Detection and delineation of squamous neoplasia with hyperspectral imaging in a mouse model of tongue carcinogenesis. *Journal of Biophotonics*, 11(3). <https://doi.org/10.1002/jbio.201700078>
- Luzzi, K. J., MacDonald, I. C., Schmidt, E. E., Kerkvliet, N., Morris, V. L., Chambers, A. F., & Groom, A. C. (1998). Multistep nature of metastatic inefficiency: Dormancy of solitary cells after successful extravasation and limited survival of early micrometastases. *American Journal of Pathology*, 153(3), 865–873. [https://doi.org/10.1016/S0002-9440\(10\)65628-3](https://doi.org/10.1016/S0002-9440(10)65628-3)
- Ma, L., Lu, G., Wang, D., Wang, X., Chen, Z. G., Muller, S., Chen, A., & Fei, B. (2017). Deep learning based classification for head and neck cancer detection with hyperspectral imaging in an animal model. *Medical Imaging 2017: Biomedical Applications in Molecular, Structural, and Functional Imaging*, 10137, 101372G. <https://doi.org/10.1117/12.2255562>
- Martineau, D., Lemberger, K., Dallaire, P. A., Lipscomb, T. P., Michel, P., & Mikaelian, I. (2002). Cancer in wildlife, a case study: Beluga in St Lawrence Estuary, Québec, Canada. *Environmental Health Perspectives*, 110(3), 285–295.
- Marx, V. (2019). Machine learning, practically speaking. *Nature Methods*, 16(6), 463–467. <https://doi.org/10.1038/s41592-019-0432-9>
- Metzger, M. J., Villalba, A., Carballal, M. J., Iglesias, D., Sherry, J., Reinisch, C., Muttray, A. F., Baldwin, S. A., & Goff, S. P. (2016). Widespread transmission of independent cancer lineages within multiple bivalve species. *Nature*, 534(7609), 705–709.
- Minakshi, M., Bharti, P., & Chellappan, S. (2017). Identifying mosquito species using smart-phone cameras. In *EuCNC 2017-European Conference on Networks and Communications* (pp. 1–6). <https://doi.org/10.1109/EuCNC.2017.7980646>
- Møller, A. P., & Mousseau, T. A. (2007). Birds prefer to breed in sites with low radioactivity in Chernobyl. *Proceedings of the Royal Society B: Biological Sciences*, 274(1616), 1443–1448. <https://doi.org/10.1098/rspb.2007.0005>
- Motta, D., Santos, A. Á. B., Winkler, I., Machado, B. A. S., Pereira, D. A. D. I., Cavalcanti, A. M., Fonseca, E. O. L., Kirchner, F., & Badaró, R. (2019). Application of convolutional neural networks for classification of adult mosquitoes in the field. *PLoS ONE*, 14(1), 1–18. <https://doi.org/10.1371/journal.pone.0210829>
- Mountrakis, G., Im, J., & Ogole, C. (2011). Support vector machines in remote sensing: A review. *ISPRS Journal of Photogrammetry and Remote Sensing*, 66(3), 247–259. <https://doi.org/10.1016/j.isprsjprs.2010.11.001>
- Myers, M. S., Landahl, J. T., Krahn, M. M., & McCain, B. B. (1991). Relationships between hepatic neoplasms and related lesions and exposure to toxic chemicals in marine fish from the U.S. west coast. *Environmental Health Perspectives*, 90, 7–15. <https://doi.org/10.2307/3430840>
- Nathan, R., Getz, W. M., Revilla, E., Holyoak, M., Kadmon, R., Saltz, D., & Smouse, P. E. (2008). A movement ecology paradigm for unifying organismal movement research. *Proceedings of the National Academy of Sciences of the United States of America*, 105(49), 19052–19059. <https://doi.org/10.1073/pnas.0800375105>
- Nicholson, B. D., Hamilton, W., O'Sullivan, J., Aveyard, P., & Hobbs, F. D. R. (2018). Weight loss as a predictor of cancer in primary care: A systematic review and meta-analysis. *British Journal of General Practice*, 68(670), e311–e322. <https://doi.org/10.3399/bjgp18X695801>
- Olden, J. D., Lawler, J. J., & Poff, N. L. (2008). Machine learning methods without Tears: A primer for ecologists. *The Quarterly Review of Biology*, 83(2), 171–193. <https://doi.org/10.1086/587826>
- Pantel, K., & Alix-Panabières, C. (2010). Circulating tumour cells in cancer patients: Challenges and perspectives. *Trends in Molecular Medicine*, 16(9), 398–406. <https://doi.org/10.1016/j.molmed.2010.07.001>
- Perrault, J. R., Stacy, N. I., Lehner, A. F., Mott, C. R., Hirsch, S., Gorham, J. C., Buchweitz, J. P., Bresette, M. J., & Walsh, C. J. (2017). Potential effects of brevetoxins and toxic elements on various health variables in Kemp's ridley (*Lepidochelys kempii*) and green (*Chelonia mydas*) sea turtles after a red tide bloom event. *Science of the Total Environment*, 605–606, 967–979. <https://doi.org/10.1016/j.scitotenv.2017.06.149>
- Perret, C., Gidoïn, C., Ujvari, B., Thomas, F., & Roche, B. (2020). Predation shapes the impact of cancer on population dynamics and the evolution of cancer resistance. *Evolutionary Applications*, 13(7), 1733–1744. <https://doi.org/10.1111/eva.12951>
- Pesavento, P. A., Agnew, D., Keel, M. K., & Woolard, K. D. (2018). Cancer in wildlife: Patterns of emergence. *Nature Reviews Cancer*, 18(10), 646–661. <https://doi.org/10.1038/s41568-018-0045-0>
- Pye, R. J., Pemberton, D., Tovar, C., Tubio, J. M. C., Dun, K. A., Fox, S., Darby, J., Hayes, D., Knowles, G. W., Kreiss, A., Siddle, H. V. T., Swift, K., Lyons, A. B., Murchison, E. P., & Woods, G. M. (2016). A second transmissible cancer in Tasmanian devils. *Proceedings of the National Academy of Sciences of the United States of America*, 113(2), 374–379. <https://doi.org/10.1073/pnas.1519691113>
- Qian, Y., Zhou, W., Yan, J., Li, W., & Han, L. (2015). Comparing machine learning classifiers for object-based land cover classification using very high resolution imagery. *Remote Sensing*, 7(1), 153–168. <https://doi.org/10.3390/rs70100153>
- Redmon, J., Divvala, S., Girshick, R., & Farhadi, A. (2015). You Only Look Once: Unified, real-time object detection, *arXiv Preprint arXiv:1506.02640*.
- Ren, J., Hacihaliloglu, I., Singer, E. A., Foran, D. J., & Qi, X. (2019). Unsupervised domain adaptation for classification of histopathology whole-slide images. *Frontiers in Bioengineering and Biotechnology*, 7(May), 1–12. <https://doi.org/10.3389/fbioe.2019.00102>
- Ren, S., He, K., & Girshick, R. (2015). Faster R-CNN: towards real-time object detection with region proposal networks. In *Advances in Neural Information Processing Systems 28 (NIPS 2015)* (pp. 91–99).
- Rodellar, J., Alférez, S., Acevedo, A., Molina, A., & Merino, A. (2018). Image processing and machine learning in the morphological analysis of blood cells. *International Journal of Laboratory Hematology*, 40, 46–53. <https://doi.org/10.1111/ijlh.12818>
- Ruiz-Aravena, M., Jones, M. E., Carver, S., Estay, S., Espejo, C., Storfer, A., & Hamede, R. K. (2018). Sex bias in ability to cope with cancer: Tasmanian devils and facial tumour disease. *Proceedings of the Royal Society B: Biological Sciences*, 285(1891), 20182239. <https://doi.org/10.1098/rspb.2018.2239>
- Salem, H., Attiya, G., & El-Fishawy, N. (2017). Classification of human cancer diseases by gene expression profiles. *Applied Soft Computing Journal*, 50, 124–134. <https://doi.org/10.1016/j.asoc.2016.11.026>
- Schlegl, T., Seeböck, P., Waldstein, S. M., Langs, G., & Schmidt-Erfurth, U. (2019). f-AnoGAN: Fast unsupervised anomaly detection with generative adversarial networks. *Medical Image Analysis*, 54, 30–44. <https://doi.org/10.1016/j.media.2019.01.010>
- Schlegl, T., Seeböck, P., Waldstein, S. M., Schmidt-Erfurth, U., & Langs, G. (2017). Unsupervised anomaly detection with generative adversarial networks to guide marker discovery. In *Lecture Notes in Computer*

- Science (Including Subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics)*, 10265 LNCS (pp. 146–147). https://doi.org/10.1007/978-3-319-59050-9_12
- Schlemmer, H.-P., Bittencourt, L. K., D'Anastasi, M., Domingues, R., Khong, P.-L., Lockhat, Z., Mueller, A., Reiser, M. F., Schilsky, R. L., & Hricak, H. (2018). Global challenges for cancer imaging. *Journal of Global Oncology*, 2018(4). <https://doi.org/10.1200/JGO.17.00036>
- Schmidt, U., Weigert, M., Broaddus, C., & Myers, G. (2018). Cell detection with star-convex polygons. In *Lecture notes in computer science (Including Subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics)*, 11071 LNCS (pp. 265–273). https://doi.org/10.1007/978-3-030-00934-2_30
- Schofield, G., Katselidis, K. A., Lilley, M. K. S., Reina, R. D., & Hays, G. C. (2017). Detecting elusive aspects of wildlife ecology using drones: New insights on the mating dynamics and operational sex ratios of sea turtles. *Functional Ecology*, 31(12), 2310–2319. <https://doi.org/10.1111/1365-2435.12930>
- Siddiqui, S. A., Salman, A., Malik, M. I., Shafait, F., Mian, A., Shortis, M. R., & Harvey, E. S. (2018). Automatic fish species classification in underwater videos: Exploiting pre-trained deep neural network models to compensate for limited labelled data. *ICES Journal of Marine Science*, 75(1), 374–389. <https://doi.org/10.1093/icesjms/fsx109>
- Sikpa, D., Fouquet, J. P., Lebel, R., Diamandis, P., Richer, M., & Lepage, M. (2019). Automated detection and quantification of breast cancer brain metastases in an animal model using democratized machine learning tools. *Scientific Reports*, 9(1), 1–8. <https://doi.org/10.1038/s41598-019-53911-x>
- Siska, V., Eriksson, A., Mehlig, B., & Manica, A. (2018). A metapopulation model of the spread of the Devil Facial Tumour Disease predicts the long term collapse of its host but not its extinction. *BioRxiv*, 347062. <https://doi.org/10.1101/347062>
- Skazina, M., Odintsova, N., Maiorova, M., Ivanova, A., Väinölä, R., & Strelkov, P. (2021). First description of a widespread *Mytilus trossulus*-derived bivalve transmissible cancer lineage in *M. trossulus* itself. *Scientific Reports*, 11(1), 5809. <https://doi.org/10.1038/s41598-021-85098-5>
- Song, Q. Z., Zhao, L., Luo, X. K., & Dou, X. C. (2017). Using deep learning for classification of lung nodules on computed tomography images. *Journal of Healthcare Engineering*, 2017, Article ID 8314740. <https://doi.org/10.1155/2017/8314740>
- Stepanski, E. J., Walker, M. S., Schwartzberg, L. S., Blakely, L. J., Ong, J. C., & Houts, A. C. (2009). The relation of trouble sleeping, depressed mood, pain, and fatigue in patients. *Journal of Clinical Sleep Medicine*, 5(2), 132–136.
- Strobl, C., Malley, J., & Gerhard, T. (2009). Characteristics of classification and regression trees, bagging and random forests. *Psychological Methods*, 14(4), 323–348.
- Swimmer, J. Y. (2006). Relationship between basking and fibropapillomatosis in captive green turtles (*Chelonia mydas*). *Chelonian Conservation and Biology*, 5(2), 305–309. [https://doi.org/10.2744/1071-8443\(2006\)5\[305:RBBAFI\]2.0.CO;2](https://doi.org/10.2744/1071-8443(2006)5[305:RBBAFI]2.0.CO;2)
- Tabak, M. A., Norouzzadeh, M. S., Wolfson, D. W., Sweeney, S. J., Vercauteren, K. C., Snow, N. P., Halseth, J. M., Di Salvo, P. A., Lewis, J. S., White, M. D., Teton, B., Beasley, J. C., Schlichting, P. E., Boughton, R. K., Wight, B., Newkirk, E. S., Ivan, J. S., Odell, E. A., Brook, R. K., ... Miller, R. S. (2019). Machine learning to classify animal species in camera trap images: Applications in ecology. *Methods in Ecology and Evolution*, 10(4), 585–590. <https://doi.org/10.1111/2041-210X.13120>
- Temitope Yekeen, S., Balogun, A. L., & Wan Yusof, K. B. (2020). A novel deep learning instance segmentation model for automated marine oil spill detection. *ISPRS Journal of Photogrammetry and Remote Sensing*, 167(July), 190–200. <https://doi.org/10.1016/j.isprsjprs.2020.07.011>
- Thessen, A. (2016). Adoption of machine learning techniques in ecology and earth science. *One Ecosystem*, 1, e8621. <https://doi.org/10.3897/oneeco.1.e8621>
- Thomas, F., Vavre, F., Tissot, T., Vittecoq, M., Giraudeau, M., Bernex, F., Misse, D., Renaud, F., Raven, N., Beckmann, C., Hamede, R., Biro, P. A., & Ujvari, B. (2018). Cancer is not (only) a senescence problem. *Trends in Cancer*, 4(3), 169–172. <https://doi.org/10.1016/j.trecan.2018.01.002>
- Valletta, J. J., Torney, C., Kings, M., Thornton, A., & Madden, J. (2017). Applications of machine learning in animal behaviour studies. *Animal Behaviour*, 124, 203–220. <https://doi.org/10.1016/j.anbehav.2016.12.005>
- Varley, A., Tyler, A., Smith, L., Dale, P., & Davies, M. (2016). Mapping the spatial distribution and activity of 226Ra at legacy sites through Machine Learning interpretation of gamma-ray spectrometry data. *Science of the Total Environment*, 545–546, 654–661. <https://doi.org/10.1016/j.scitotenv.2015.10.112>
- Vittecoq, M., Ducasse, H., Arnal, A., Møller, A. P., Ujvari, B., Jacqueline, C. B., Tissot, T., Missé, D., Bernex, F., Pirot, N., Lemberger, K., Abadie, J., Labrut, S., Bonhomme, F., Renaud, F., Roche, B., & Thomas, F. (2015). Animal behaviour and cancer. *Animal Behaviour*, 101, 19–26. <https://doi.org/10.1016/j.anbehav.2014.12.001>
- Vittecoq, M., Roche, B., Daoust, S. P., Ducasse, H., Missé, D., Abadie, J., Labrut, S., Renaud, F., Gauthier-Clerc, M., & Thomas, F. (2013). Cancer: A missing link in ecosystem functioning? *Trends in Ecology & Evolution*, 28(11), 628–635. <https://doi.org/10.1016/j.tree.2013.07.005>
- Wagstaff, K. (2012). Machine learning that matters. In J. Langford, & J. Pineau (Eds.), *Proceedings of 29th International Conference on Machine Learning (ICML-12)* (pp. 529–536).
- Wäldchen, J., & Mäder, P. (2018). Machine learning for image based species identification. *Methods in Ecology and Evolution*, 9(11), 2216–2225. <https://doi.org/10.1111/2041-210X.13075>
- Wang, S., Zhou, Y., Qin, X., Nair, S., Huang, X., & Liu, Y. (2020). Label-free detection of rare circulating tumor cells by image analysis and machine learning. *Scientific Reports*, 10(1), 1–10. <https://doi.org/10.1038/s41598-020-69056-1>
- Wardeh, M., Sharkey, K. J., & Baylis, M. (2020). Integration of shared-pathogen networks and machine learning reveals the key aspects of zoonoses and predicts mammalian reservoirs. *Proceedings of the Royal Society B: Biological Sciences*, 287(1920), 20192882. <https://doi.org/10.1098/rspb.2019.2882>
- Weinstein, B. G. (2018). A computer vision for animal ecology. *Journal of Animal Ecology*, 87(3), 533–545. <https://doi.org/10.1111/1365-2656.12780>
- Wijeyakulasuriya, D. A., Eisenhauer, E. W., Shaby, B. A., & Hanks, E. M. (2020). Machine learning for modeling animal movement. *PLoS ONE*, 15(7 July), 1–30. <https://doi.org/10.1371/journal.pone.0235750>
- Williams, G. J., Aebys, G. S., Cowie, R. O. M., & Davy, S. K. (2010). Predictive modeling of coral disease distribution within a reef system. *PLoS ONE*, 5(2), 1–10. <https://doi.org/10.1371/journal.pone.0009264>
- Wright, T., Brisson, B. A., Wood, G. A., Oblak, M., Mutsaers, A. J., Sabine, V., Skowronski, K., Belanger, C., Tiessen, A., & Bienze, D. (2019). Flow cytometric detection of circulating osteosarcoma cells in dogs. *Cytometry Part A*, 95(9), 997–1007. <https://doi.org/10.1002/cyto.a.23847>
- Xia, C., Hu, Y. I., Ward, M. P., Lynn, H., Li, S. I., Zhang, J., Hu, J., Xiao, S., Lu, C., Li, S., Liu, Y., & Zhang, Z. (2019). Identification of high-risk habitats of oncomelania hupensis, the intermediate host of schistosoma japonicum in the poyang lake region, China: A spatial and ecological analysis. *PLoS Neglected Tropical Diseases*, 13(6), 1–16. <https://doi.org/10.1371/journal.pntd.0007386>
- Xue, X., Lin, Y. I., Zheng, Q., Wang, K. E., Zhang, J., Deng, J., Abubakar, G. A., & Gan, M. (2020). Mapping the fine-scale spatial pattern of artificial light pollution at night in urban environments from the perspective of bird habitats. *Science of the Total Environment*, 702, 134725. <https://doi.org/10.1016/j.scitotenv.2019.134725>
- Yu, H., Deng, J., Nathan, R., Kröschel, M., Pekarsky, S., Li, G., & Klaassen, M. (2021). An evaluation of machine learning classifiers for

- next-generation, continuous-ethogram smart trackers. *Movement Ecology*, 9(1), 1–15. <https://doi.org/10.1186/s40462-021-00245-x>
- Zapata, L., Chalco, L., Aguilar, L., Pimbosa, E., Ramírez-Morales, I., Hidalgo, J., Yandún, M., Arias-Flores, H., & Guevara, C. (2020). Detection of Cutaneous Tumors in Dogs Using Deep Learning Techniques. In T. Ahram (Ed.), *Advances in artificial intelligence, software and systems engineering* (pp. 83–91). Springer International Publishing.
- Zou, Q., Xie, S., Lin, Z., Wu, M., & Ju, Y. (2016). Finding the best classification threshold in imbalanced classification. *Big Data Research*, 5, 2–8. <https://doi.org/10.1016/j.bdr.2015.12.001>

How to cite this article: Dujon, A. M., Vittecoq, M., Bramwell, G., Thomas, F., & Ujvari, B. (2021). Machine learning is a powerful tool to study the effect of cancer on species and ecosystems. *Methods in Ecology and Evolution*, 12, 2310–2323. <https://doi.org/10.1111/2041-210X.13703>