

25TH ANNIVERSARY PAPER

# An Outlook for Deep Learning in Ecosystem Science

George L. W. Perry,<sup>1\*</sup> Rupert Seidl,<sup>2,3</sup> André M. Bellvé,<sup>1</sup> and Werner Rammer<sup>2</sup>

<sup>1</sup>*School of Environment, University of Auckland, Auckland 1142, New Zealand;* <sup>2</sup>*Ecosystem Dynamics and Forest Management Group, School of Life Sciences, Technical University of Munich, Freising, Germany;* <sup>3</sup>*Berchtesgaden National Park, Berchtesgaden, Germany*

## ABSTRACT

Rapid advances in hardware and software, accompanied by public- and private-sector investment, have led to a new generation of data-driven computational tools. Recently, there has been a particular focus on deep learning—a class of machine learning algorithms that uses deep neural networks to identify patterns in large and heterogeneous datasets. These developments have been accompanied by both hype and scepticism by ecologists and others. This review describes the context in which deep learning methods have emerged, the deep learning methods most relevant to ecosystem ecologists, and some of the problem domains they have been applied to. Deep learning methods have high predictive performance in a range of ecological contexts, leveraging the large data resources now available. Furthermore, deep learning tools offer ecosystem ecologists new ways to learn about ecosystem dynamics. In particular, recent advances in interpretable machine learning and in developing hybrid approaches combining deep learning and mechanistic models provide a bridge between

pure prediction and causal explanation. We conclude by looking at the opportunities that deep learning tools offer ecosystem ecologists and assess the challenges in interpretability that deep learning applications pose.

**Key words:** artificial intelligence; deep learning; machine learning; neural networks; interpretability; ecosystem ecology; environmental science.

## Glossary

- Activation function** The activation function determines the output value of a neuron in a neural network as a function of the sum of the weighted inputs it receives (see also ‘weights’)
- Active learning** A method in which model training has an interactive component where after some training the model periodically supplies the users with cases to label—the goal is to minimise the amount of data required for adequate training
- Artificial Intelligence (AI)** The suite of computational tools and efforts that seek to mimic human intelligence and information processing

Received 26 June 2022; accepted 1 September 2022

**Supplementary Information:** The online version contains supplementary material available at <https://doi.org/10.1007/s10021-022-00789-y>.

**Authors Contributions:** GP wrote the first draft of the manuscript, and all authors made substantial contributions to subsequent revisions. GP and AB designed the bibliometric analysis, and AB conducted it.

\*Corresponding author; e-mail: george.perry@auckland.ac.nz

Artificial Neural Networks (ANN)	Artificial neural networks are computational systems designed to imitate how a human brain processes information. In these models, ‘neurons’ successively transform information to predict some outcome	Error function	Also known as a loss function, the error function quantifies the deviation of a network’s prediction from the ground true value. This error is minimised during model training
Backpropagation	Backpropagation is the minimisation of some loss function to optimise a model’s parameterisation—in the case of an ANN this involves tuning the weights. For DNN the process works backward from the final fully connected layer	Feature extraction	The conversion of (raw) input data into (simpler) representations to make classifications or predictions without losing key characteristics of the original data. Feature extraction often results in better DL model performance than using the raw data
Convolutional Layers	The core component of a CNN architecture, the convolutional layers apply the same set of learned filters over the entire image to extract the features that will be used for classification and prediction	Feature selection	The process of input variable selection (synonymous with variable selection in classic frequentist regression models). The reduction of input parameters can aid in both computational costs and spurious associations
Convolutional Neural Networks (CNN)	A form of specialised deep neural network typically used for image classification. CNNs are made of three types of layers: convolutional, pooling and fully connected layers	Feed-forward Neural Networks (FNN)	An architecture of artificial neural network in which information is fed from the input layer neurons to the hidden layer neurons before being transferred to output layer neurons. Information flow is in one direction only (input layer → hidden layers → output layer)
Data augmentation	A technique used to synthesise new training data by modifying existing data (for example, image rotation, image cropping, and so on). This technique can be used to mitigate against overfitting when training a DL model with small training sets	Fully connected layers	In a fully connected layer of an ANN, all neurons are connected with all other neurons in the preceding and succeeding layers
Deep learning (DL)	Deep learning is the component of machine learning that uses neural network algorithms with multiple hidden layers (the multiple layers make them ‘deep’)	Hidden layer(s)	The layer(s) between input and output layers containing neurons. These neurons receive inputs which are weighted and then produce outputs based on an activation function
Edge computing	A technological framework that distributes computing so as to process data at (or as close as possible to) source to limit bandwidth usage, energy consumption, and so on	Input Layer	The layer in an ANN which receives the initial raw data, processes it and passes it onto the hidden layers. The input layer is the first step in an artificial neural network

Layers	Neural networks are made up of a series of layers which receive the input information (input layer), process it through a series of (hidden) layers before making predictions (output layer). There are typically three categories of layers in neural networks; input layers, hidden layers and output layers, although their organisation will vary depending on the ANN architecture	Pooling layers	The pooling layer in a CNN aggregates information by merging the results of multiple CNN filters
Long short-term memory (LSTM)	An architecture designed to overcome some technical problems with training RNN models. In place of neurons, these models use memory blocks that are connected between layers. Each block has a memory of recent sequences and a gate that controls its state and the information it outputs; this architecture allows the information flow from a block to be conditional on its state	Recurrent Neural Network (RNN)	An ANN that can represent auto-correlation between data points by incorporating dependencies between observations. This architecture makes RNNs particularly useful for predicting time-series data
Machine Learning (ML)	Machine learning is a subset of AI that develops algorithms designed to iteratively learn (for example, identify patterns) from data	Saliency maps	Heatmaps developed to highlight the most important portions of an image in a DL model (usually a CNN) making a prediction—they are a tool designed to improve DL model interpretability
Neurons/Nodes	Each layer in a neural network is comprised of a series of neurons, each of which is a mathematical operation. These neurons apply the operation to incoming data, multiply it by a weight, and pass the resulting value through an activation function to other neurons in the network.	Supervised training	The practice of providing the ANN with data that are ‘labelled’ in some way (for example, wildlife imagery with the species in the image tagged). This process enables training a model for a particular predictive task and then assessing its (predictive) performance
Output Layer	The final layer in an artificial neural network where the information that was processed by the hidden layers is reformulated to create the desired predictions. The neurons in this layer also have their own weights that are applied to aid in the derivation of the prediction	Testing data	Data used to test a model’s performance for a given task; often a subset of all data available and not used in model training
		Training data	Data used to train a model for a particular task. These data are typically held separate to the testing data to prevent overfitting
		Transfer learning	The practice of using knowledge gained from solving one problem in a separate but related problem. In the context of DL models, this is applying a model trained in one context in a new setting
		Unsupervised training	The practice of training a model with unlabelled input data; clustering algorithms are a well-known example of this approach

**Weights** The weights in an ANN model control the information flowing from a node (in some ways analogous to the slope in a regression model). The weights are combined for all nodes in a layer in an activation function which determines how information is passed out of a layer

---

## INTRODUCTION

Although their origins lie in the 1940s (Goodfellow and others 2016), the past decade has seen rapidly growing application of tools associated with artificial intelligence (AI), machine learning (ML) and deep learning (DL) across the sciences. In the Google Scholar Metrics report for 2020, the most cited papers across all subject areas are dominated by those in the field of AI, including three of the top five in *Nature* (the other two being in genetics).<sup>1</sup> This trend reflects the rapid progress and growing importance of AI methods and tools in many fields, including computer vision, sound classification, natural language processing, gaming, and robotics. The increasing use of AI, ML and DL methods has been driven by a combination of technical developments in the algorithms themselves, the availability of large data sets, rising computational power (including cloud-based services, GPU-optimised code, specialist processor units, edge computing), and accessible open-source frameworks for their implementation. It has also been driven by massive funding from the private and public sectors, partly due to (sometimes hyperbolic) assessments of the opportunities offered by AI. Alongside these advances, there has been growing concern over AI's ethical and privacy challenges. While developments in hardware, software and knowledge offer potentially transformative opportunities for ecologists (for example, powerful tools to work with new data sources such as images, audio and language), as with many new technologies, they have been over-hyped, which has led to some cynicism as to what they offer. As we will discuss, DL methods—while certainly not

the be-all and end-all of analysis methods—have great potential to advance ecosystem ecology.

Even against the background of the exponential growth in the scientific literature (Wang and Barabási 2021), there has been an explosion of publications discussing or using AI and deep learning in the environmental sciences since the mid-2000s (Figure 1a). Analysis of the keywords in a corpus of papers considering artificial intelligence (see Supplementary Material) identified three broad topic areas (Figure 1b): (i) environmental modelling and forecasting (for example, time-series analysis of water or air quality), (ii) automated image detection and classification (for example, identification of species in wildlife camera traps) and (iii) remote sensing and landscape classification (for example, image classification for forest disturbance detection from satellite data). This brief analysis underscores that ecologists increasingly apply DL approaches in a range of problem domains.

Current perspectives on DL among ecologists range from 'DL is a universal panacea' to 'DL is an inscrutable black box' to 'DL methods are an over-hyped fad'. In this review, we seek to provide a realistic perspective of how to best capitalise on the investments by the public and private sectors in these technologies and leverage those developments to foster new avenues for ecosystem research. Our review is not intended as a 'how to' primer, nor is it aimed at experts in DL methods. Likewise, it is not a comprehensive evaluation of every potential or realised application of DL in the context of ecosystem ecology. Instead, it is intended to introduce what DL and associated methods might offer ecosystem ecology and some of the challenges these applications pose. We briefly describe the neural networks that underpin DL, and consider their application in three contrasting problem domains, before concluding with how ecosystem ecologists might best exploit these new approaches.

## Deep Learning Algorithms

Deep learning relies on artificial neural networks (ANN), which are loosely modelled on the brain with artificial neurons (nodes) connected so they can communicate (analogous to synaptic connections). Deep neural networks (DNN) have become widely used during the past decade but descend from simpler artificial neural networks devised in the 1950s and 1960s (Figure 2; Goodfellow and others 2016; Razavi 2021). While these early networks mainly used one hidden layer of nodes (Figure 2), DNNs have many of them; hence, the moniker 'deep' (see Glossary for expanded defini-

---

<sup>1</sup> <https://www.natureindex.com/news-blog/google-scholar-reveals-most-influential-papers-research-citations-twenty-twenty>.

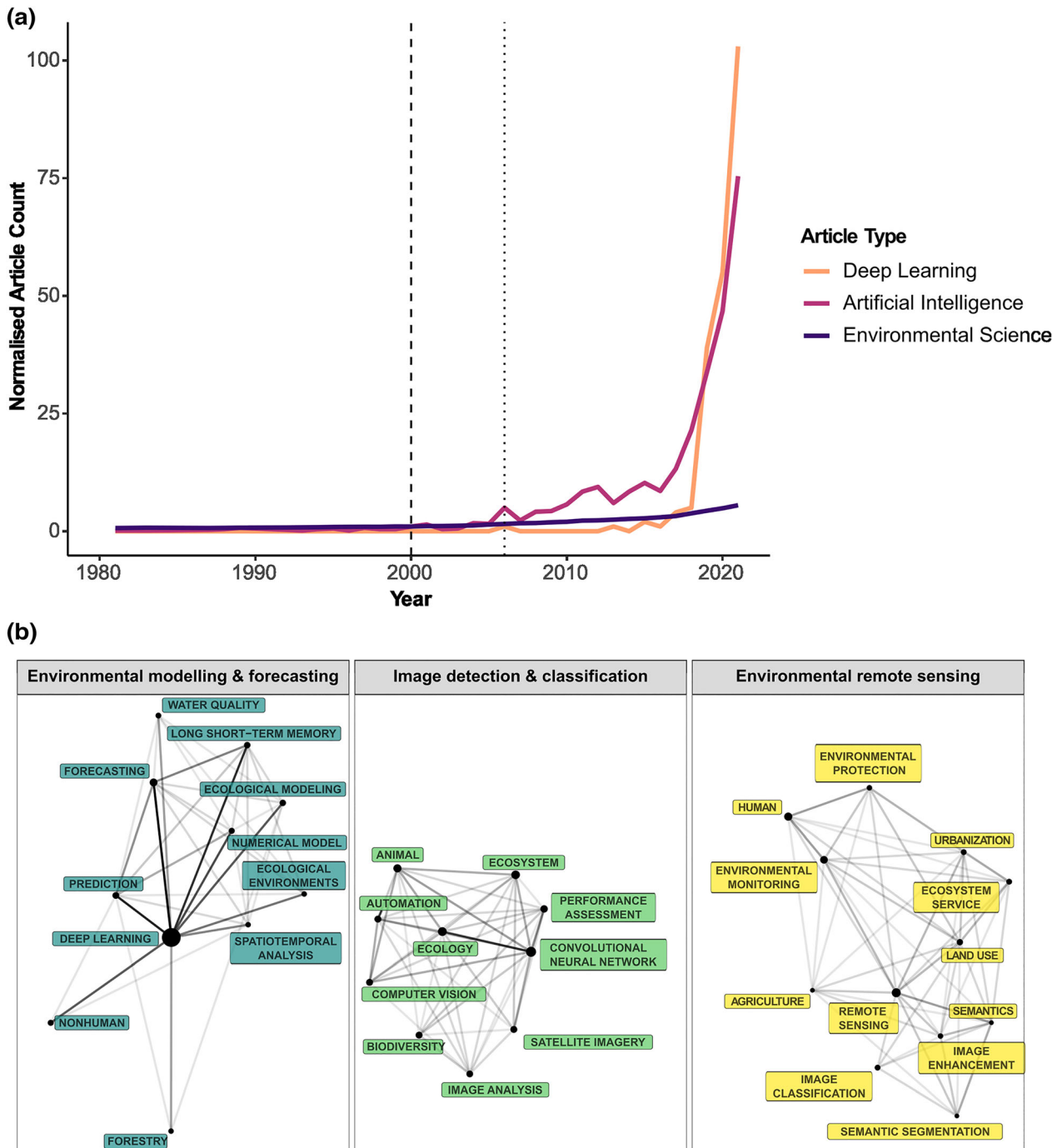
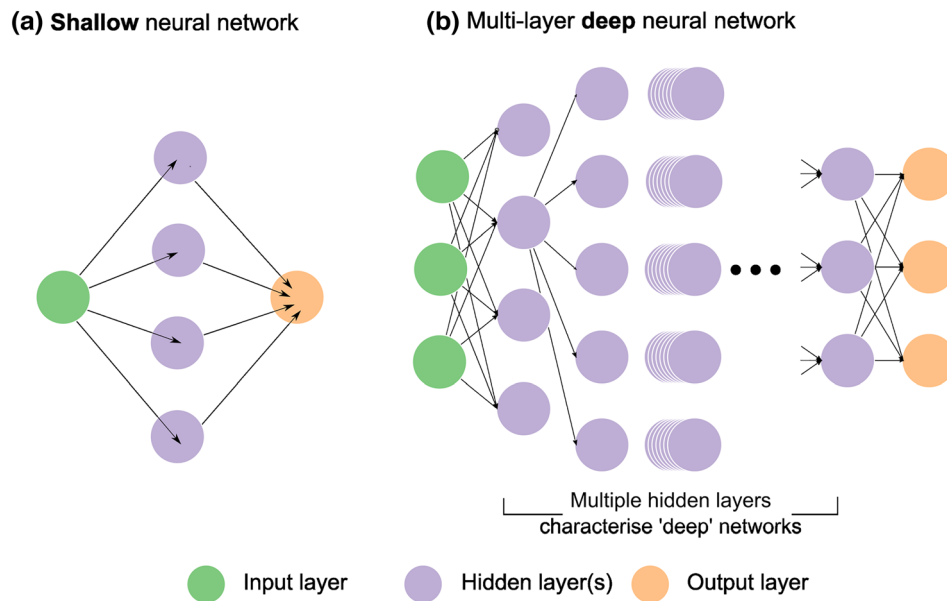


Figure 1. The rise of AI and DL in the scientific literature: **a** normalised publication rate in articles in environmental science overall, artificial intelligence, and deep learning from 1971 to 2021; and **b** focal areas of publications in publications in environmental science using deep learning tools based on keyword co-occurrence analysis. In **b** the groups are identified by maximising the density of connections between vs. across groups (that is, modularity; R command `igraph::group_fast_greedy`), with the size of the nodes (black points) proportional to the frequency that a keyword appeared together and the weight of the edges (lines between nodes) is proportional to the strength of the link (frequency of co-occurrences). The normalisation in **a** is the publication rate in a given year relative to 2000 for artificial intelligence and environmental science, while deep learning is normalised to 2006, as this was the earliest paper published with deep learning in the environmental science discipline. See Supplementary Material for details.



**Figure 2.** A schematic of neural network architecture for **a** a shallow neural network with just one hidden layer, and **b** a deep feed-forward neural network with multiple hidden layers (here depicted as uniform and fully connected layers with not all links shown).

tions). Depending on the application, the inputs to a DNN (Figure 2) can be diverse, including pixels in an image, words in a sentence, data points in a time series, and can be mixed in type (qualitative, categorical, quantitative). Similarly, the type of outputs can vary with classification (the network determines for the given inputs one of a pre-defined set of classes) and regression (the network determines a single numeric value from the input data).

DNNs vary in their architecture, that is the details of the ‘wiring’ of the nodes. The most straightforward architecture is the feed-forward neural network (Figure 2 right), in which raw data are successively transformed into more abstract representations until some output is produced (for example, identification of a species from an audio recording). In this architecture, nodes (the neurons) are fully connected between but not within layers. In a simple (feed-forward) ANN, input data are transformed by a sequence of nodes in a ‘hidden’ layer to generate output (Figure 2). In a shallow ANN, the single hidden layer will consist of a series of nodes that can transform data using a sigmoidal function, based on the fact that any data transformation can be achieved using a stack of sigmoidal functions and a linear transform (Goodfellow and others 2016; Borowiec and others 2022). Although in theory given sufficient nodes a shallow neural network can apply any transformation, it is more efficient to use multiple layers

than one single enormous one (Razavi 2021). Razavi (2021) provides an intuitive and thorough geometric explanation of these ideas, which he calls the concept of ‘depth’ (p. 4). In DNN, the multiple layers have different purposes; for example, in a CNN different layers may apply convolution kernels to extract key features from an image and pooling layers to generalise (down sample) these features. The architecture of deep neural networks is such that layers go from general to specific, with the last layer fully connected and producing the output. Thus, as Yosinski and others (2014) discuss, when DNNs are trained to classify images, the first layer tends to identify similar high-level features (“Gabor filters or colour blobs”, p. 1) irrespective of the image type. Each node is characterised by an activation function that defines how the values from incoming connections are combined and forwarded to the next layer of nodes. In a fully connected network, each node is connected to all nodes in the following layer via variable weights that are learned and hence represent the relationships between variables. The DNN ‘learns’ by optimising the connection weights in the network to minimise the prediction error (Olden and others 2008; LeCun and others 2015).

Learning is most commonly performed using a backpropagation algorithm, where an error function is minimised iteratively across observations by updating the weights to decrease their contribution to the overall error (Olden and others 2008). In



supervised training (that is, the output variable [response] is labelled, and the algorithm trained to predict the label), the goal is to minimise the difference between the predicted and observed values. After a DNN is trained, it can be used to predict outcomes based on new data. Usually, the output is directly used for classification (for example, what animal is in the picture?) or regression (for example, what is the predicted value of a time series?), but DNNs can be the (core) part of more complex toolchains. Image segmentation (a.k.a. image semantics, assigning each pixel in an image to a class) or object detection (detecting the type and position of multiple objects in an image) are examples derived from computer vision that have become regularly used in ecology.

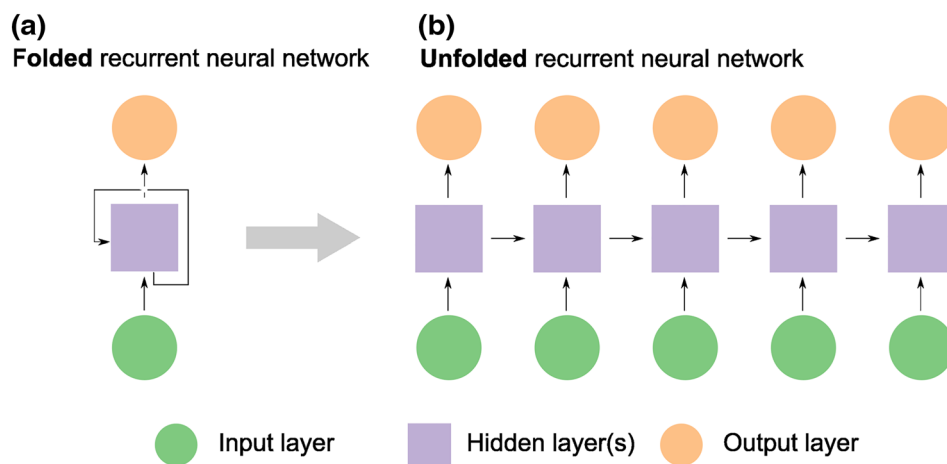
As noted previously, there are many ways to wire a DNN's nodes (that is, the architecture of the DNN). Ecologists have most frequently used convolutional neural networks (Brodrick and others 2019; Christin and others 2019; Borowiec and others 2022), as they are particularly well-suited for image and audio processing. In a convolutional neural network (CNN), the hidden layers comprise convolution layers (hence the name), pooling layers, and fully connected layers designed for different components of image recognition (feature extraction, downscaling, and integration); Rawat and Wang (2017) review the design and application of CNNs for image classification. Another architecture of potential importance for ecosystem ecologists is recurrent neural networks (RNN; Figure 3). RNNs process sequences and keep a 'memory' of past data by feeding the output of a layer back into that same layer (hence 'recurrent', Figure 3a). RNNs can be imagined as a sequence of

neural networks feeding each other by sharing parameter information (the unfolded neural network in Figure 3b). The length of time at which a previous network state is influential will depend on changes in weights during training; thus, in principle, RNNs can deal with short- and long-term memory effects or dependencies (Goodfellow and others 2016). This architecture is well-suited to time-series applications, such as forecasting hydrological and meteorological conditions (Rahmani and others 2021; Zhi and others 2021). Because of their ability to deal with sequential data, RNNs offer a route to the near-time ecological forecasting advocated by Dietze and others (2018).

Machine learning engineers are continuously refining existing and devising new DNN architectures. For example, the "transformer" architecture (Vaswani and others 2017) initially developed in the language domain is now increasingly used for image processing tasks (Chen and others 2021a). While this complexity could seem overwhelming for ecologists, deep learning software packages are increasingly available that hide most of the technical complexity and can be used from well-known computing platforms such as R, Python, and Julia.

### Changes in the Data Landscape

Machine learning and deep learning methods have emerged in an era of large datasets (potentially comprising  $> 1 \times 10^9$  items; Goodfellow and others 2016). This emergence is crucial for DL because these methods thrive disproportionately on big data compared to classical statistical approaches. DL has the potential to leverage the information hidden in such large datasets to answer ecological questions in new ways. Thus, any discussion of DL necessi-



**Figure 3.** Schematic architecture of a recurrent neural network. The architecture is shown in a folded and unfolded view (**a** and **b**, respectively).

tates considering two inter-related trends: big data and born-digital data. Big data are characterised by the three Vs: volume, velocity, and variety (LaDeau and others 2017). Volume relates to the fact that we have unprecedented amounts of data available (although ‘unprecedented’ is context-dependent and of itself unremarkable), velocity is a function of the rapidity of data generation, sometimes happening in real-time, and variety means data are heterogeneous in form and curation. Where do these data come from? Increasingly, data collection is automated using devices that remotely measure the environment, including camera traps, satellite platforms, unmanned aerial vehicles (UAVs; drones), automated audio recording devices (continuously monitoring sondes in freshwater and marine ecosystems), and simulation models (Kays and others 2020; Keitt and Abelson 2021). In some countries, these data are openly available and collected via large coordinated research programmes such as NEON (USA; Keller and others 2008) or TERN (Australia; Cleverly and others 2019). These data streams often consist of images, audio, video or unstructured texts, which are well suited for DNNs but challenging to use with traditional statistical methods. An additional source of big data is citizen science, whether in collecting information or labelling massive datasets. This diversity of sources gives rise to the fourth ‘v’, veracity (their variable uncertainties), which is crucial to understand for these data to be used effectively (Farley and others 2018). Reconciling the various types and scales of data available to ecologists is a fundamental challenge in effectively leveraging data-led methods. DL methods are excellent tools to address this broader challenge. A particular challenge for DL methods is their demand for large amounts of accurately labelled data for supervised learning; we will return to this problem later.

### The Use of DL in Ecosystem Ecology

Ecosystem ecology is the study of the dynamics of energy and matter in ecosystems, resulting from the interactions of abiotic and biotic components of such systems and occurring across multiple spatial and temporal scales. As the publications in *Ecosystems* would attest, the field has a broad remit and interfaces with nearly every other sub-discipline of ecology. To illustrate the range of applications of DL in ecosystem ecology, we will consider three broad areas: analysis of data describing energy and matter fluxes, image processing and analysis, and integration with earth system and ecosystem models. We have drawn on case studies that align with

fundamental questions of ecosystem ecology, yet in many cases, these are allied with other components of ecology. Likewise, many of the opportunities and challenges associated with using DL are not domain-specific and encompass the use of these tools across subfields of ecology (for example, the potential of large-scale text analysis and automated translation to help alleviate biases in literature syntheses).

### PROBLEM DOMAIN 1: SYNTHESIS AND PREDICTION OF MASSIVE DATA DESCRIBING ECOSYSTEM FLUXES

Global networks, such as FluxNet and automated hydrological and meteorological stations, yield vast amounts of high-resolution information describing ecosystem fluxes (Baldocchi 2020). Deep learning methods have been applied to these data to predict temporal dynamics and to assess how they might be affected by global change. Recurrent neural networks and their relatives, such as the long short-term memory (LSTM) model (a variant of RNNs), are well-suited to modelling temporal data and have begun to be used to model earth system dynamics. Kraft and others (2019) developed RNNs to predict the normalised difference vegetation index (NDVI) based on climate data, land cover and soil information using an LSTM architecture. Their models demonstrated that including memory (past data) improved model performance in both global and biome-specific models. While the gains in performance varied between biomes they were somewhat predictable from a biome’s position in climate space. For example, memory effects seem stronger in sub-tropical regions where seasonal effects are less important than sporadic climate events (for example, interspersed wet and dry periods, see also Hansen and others 2022). The strength of memory effects also varies through time in different biomes (for example, it is strong in spring in contexts where meltwater is important). Similarly, Zhi and others (2021) used an LSTM model to predict dissolved oxygen content in catchments across the conterminous USA. Because dissolved oxygen is a vital indicator of the health of freshwater ecosystems, there is a need to develop models that are transferrable to sites where data are lacking. Zhi and others (2021) trained their model on measurements of dissolved oxygen concentrations at more than 200 sites spanning 1980–2014 (minimum of  $n = 10$  points) alongside high-quality daily meteorological data and a suite of variables characterising watershed conditions. The models captured the seasonal dynamics of dissolved oxygen, although the predic-



tions were damped at some sites. Zhi and others (2021) comment that the model performance is affected by a lack of data at dissolved oxygen extremes and heterogeneous data availability. Similar methods have been used to predict critical components of the earth system at a global scale. For example, Besnard and others (2019) implemented an RNN to predict net ecosystem CO<sub>2</sub> exchange at forest sites across the globe, using a range of data for training (remotely sensed data, down-scaled climate information and eddy covariance flux information). Their model captured broad seasonal and inter-site trends but did not adequately predict extreme conditions. Besnard and others (2019) considered that this failure to capture anomalies adequately could be explained by several issues, including missing data in the remote-sensing time series and the temporal resolution and spatial content of the information. Issues of data scarcity (including labelled data) and patchiness are a recurring challenge for data-hungry models such as DNNs.

## PROBLEM DOMAIN 2: INTERROGATING IMAGE DATA

### Object Identification and Labelling

Rapid developments in computer vision have made image analysis and processing a frequent applica-

tion for DL in the environmental sciences (Figure 1b). DL-informed image processing has been used in many ecological contexts, including (i) identifying wildlife species in camera trap data, (ii) the extraction of multidimensional whole-organism phenotypic information ('phenomics'), (iii) mapping disturbance events (for example, fire and floods), and (iv) tracking organism movement. The almost archetypal application of DL in ecology has been to extract taxonomic information from imagery. In a pioneering study, Norouzzadeh and others (2018) demonstrated the ability of DL methods to identify wildlife species in motion-activated wildlife camera imagery. They trained nine DL architectures (for comparison) to detect and identify species in the Serengeti Snapshot database, which contains 3.2 million images (Swanson and others 2015). The model (Figure 4) approached or exceeded the accuracy of human volunteers, with potentially enormous (up to 99%) timesaving. For example, their model accurately identified the 75% of images not containing an organism, which considerably reduces the number of images requiring manual assessment.

DL models have been used to characterise vegetation structure and to identify and predict disturbances in forest landscapes, primarily via the

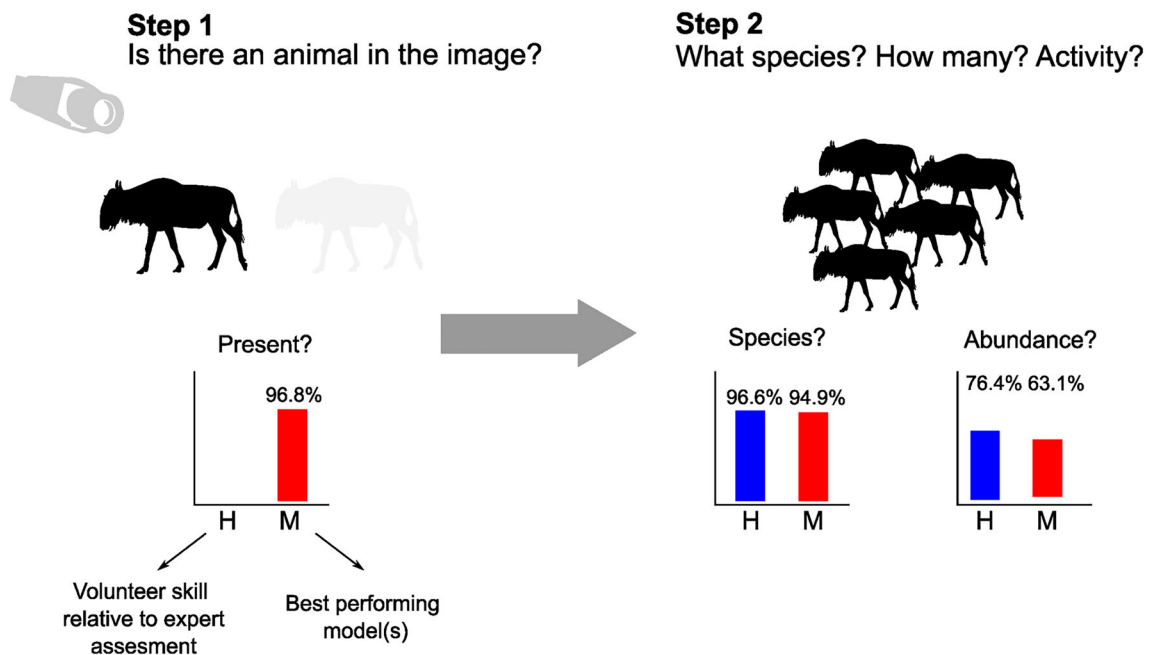


Figure 4. Schematic overview of the two-stage workflow used by Norouzzadeh and others (2018). In both steps, nine DL architectures were trained, with the best performing retained for prediction. For step 1, this was a single model (a CNN), and in step 2 it was a model ensemble. The graphs summarise model (M) vs expert (H) accuracy. There is no bar for 'H' in the 'Present?' graph as human expertise on this task was not assessed. Wildebeest icon by Lukasiniho under CC3 license (<http://creativecommons.org/licenses/by-nc-sa/3.0/>).

analysis of remotely sensed imagery; again, convolutional neural networks are the leading DL architecture used in this context. Using aerial imagery, Rammer and Seidl (2019b) trained a CNN to predict bark beetle outbreaks in a German national park within individual years and along a 23-year time series. The network was trained to predict whether a single focal cell ( $30 \times 30$  m) will be disturbed in the next year based on the average climate conditions, the spatial pattern of hosts and current disturbance in a  $600 \times 600$  m window around the focal cell. Their CNN outperformed a number of other machine learning methods and did so without the inclusion of meteorological data, on the grounds that such data are often scarce or unavailable. These applications are not limited to landscape-level dynamics. Kattenborn and others (2020) trained CNNs using UAV imagery to identify individual tree species cover in forests, estimate plant cover in a glacial vegetation succession, and identify invasion dynamics (first two cases in New Zealand, the third in Chile). Their models performed well, but they suggest important trade-offs between the accuracy and the spatial resolution of the predictions. DL models have been implemented at still finer scales to identify insects (Valan and others 2019) and pollen grains (Daood and others 2016; Olsson and others 2021). In short, DL methods are versatile, accurate, and efficient for image processing tasks; the application of these methods to ecological questions will likely continue to grow.

### Beyond Labels: Measuring Functional Traits and Behaviour

Phenotypic variation is linked to a range of ecosystem properties and functions. Studies of variation in phenotype over large spatial extents can address macroecological questions and changes over time assess how morphology tracks environmental changes (for example, body-size shifts under climate change). Manually extracting high volumes of multidimensional phenotypic data is time-consuming; hence, there is considerable interest in leveraging advances in computer vision and DL methods to facilitate this process (Lürig and others 2021).

As described previously, citizen science efforts have led to the collection of large bodies of data, especially labelled images. Schiller and others (2021) used a CNN trained using trait information from the open TRY database (Kattge and others 2020) to estimate six plant functional traits from plant images stored in the iNaturalist database.

They explored: (i) how the inclusion of intraspecific variation in traits and bioclimatic information influenced model performance and (ii) the potential for a CNN to predict traits indirectly using covariance structures (for example, leaf shape, which is apparent in the image, may predict elemental concentration in tissues). If a model can make accurate indirect trait predictions this would enable more easily measured (or cheaper) parameters to act as surrogates for more difficult ones. Schiller and others' (2021) best performing models had normalised absolute mean errors in the range of 8–15% ( $r^2 = 0.16$ – $0.58$ ) with predictions better for leaf form than tissue-related traits (that is, directly vs. indirectly measured). Similarly, Weeks and others (2022) developed a DL-based workflow to identify bones in images of bird skeletons in museum collections and measured 11 skeletal traits (the Skelevision project: <https://skelevision.net/>). This process involves detecting the bones of interest in an image (image segmentation) and then measuring them through a multistage process that used DL models to identify bones in images and to measure the characteristics of interest. Weeks and others (2022) commented that an advantage of the method was that it did not damage specimens. The accuracy of bone detection in the models depended on the morphological element; however, classification and skeletal measurement were accurate and repeatable, with only one trait showing any phylogenetic signal (for example, bias varies across taxa). Weeks and others (2022) emphasise that a critical advantage of their workflow is that it is easy to generate data describing new traits given the low annotation requirements. In short, there seems little doubt that there are many opportunities for trait-based ecology to benefit from the integration of computer vision and DL.

Data about movement can provide information about the behavioural component of phenomics (Lürig and others 2021). DL can be used to detect objects (that is, animals) in video data and track them, as well as classify such data into states potentially associated with different behaviours. These workflows involve object detection and identifying key points on the body (for pose) or tracking the objects' movement. Software toolkits have been developed that integrate computer vision and DL models to detect individuals and estimate their pose (Graving and others 2019) and movement (Walter and Couzin 2021). For example, Lopez-Marcano and others (2021) describe a workflow for detecting and tracking individual fish (bream) in video imagery. They used a CNN to identify the fish (based on a training set of 8700

annotated images) and tested three object tracking algorithms. The workflow efficiently identified and tracked individual fish and, as with other applications leveraging DL, allowed data to be collected and analysed at a scope not otherwise possible. Such applications have significance for tracking animal movement, which underpins ecosystem functions such as biogeochemical cycling and seed dispersal, and may also inform conservation activities such as identifying individuals of threatened species (Tuia and others 2022).

### PROBLEM DOMAIN 3: MODELLING ECOSYSTEM DYNAMICS

#### Hybrid Earth System and Ecosystem models

The incorporation of DL into process-based earth system models to form ‘hybrid’ model platforms is a very active research frontier (Reichstein and others 2019; Irrgang and others 2021). In such models, some system components and processes are simulated using data-driven representations and others using more mechanistic/process-based approaches. The advantage of this hybrid architecture is that it can leverage the physical consistency of process models with the data-driven performance of deep learning models (Reichstein and others 2019). There are several rationales for incorporating a DL component into ecosystem models (Reichstein and others 2019; Irrgang and others 2021): (i) to improve the estimation or upscaling of uncertain parameters, (ii) as plug-in components to replace physical models or model components, (iii) to test models by helping identify errors, and (iv) to emulate computationally expensive physical models (that is acting as meta-models).

Hybrid mechanistic-DL models have begun to be implemented to predict ecosystem properties, including evaporation (Koppa and others 2022), evapotranspiration (Chen and others 2021b), lake temperature (Read and others 2019), and snowpack distribution (Xu and others 2022). For example, Chen and others (2021b) implemented a hybrid physical-DL framework to predict daily ecosystem respiration and evapotranspiration in the western USA. Their approach combined high-resolution eddy covariance and meteorological data with land surface information (NDVI via remote sensing) to support physical and DL models (a LSTM) of evapotranspiration and ecosystem respiration. They tested the model at local (individual FLUXNET sites) and ecoregion scales (model transferability within ecoregions). At the site scale,

their model successfully captured long-term trends in evapotranspiration and ecosystem respiration; however, performance was less adequate when predicting short-term fluctuations, especially during summer extremes. Tests at the eco-region scale were also successful, although there were some issues in predicting summer extremes, demonstrating the ability of these hybrid models to predict unmeasured locations or those where data are missing. In general, Chen and others (2021b) note that the hybrid model performed well and the architecture should be extendable to other biogeochemical cycles. However, they highlight some uncertainties arising from feature selection, capturing extremes (the poorer short-term performance in summer was attributed to a lack of extremes in the training data; the earlier example of Zhi and others (2021) suffered from similar problems), resolution of meteorological information especially in mountainous terrain, issues inherent in remote sensing (for example, cloud cover), and error propagation within and between the components of the hybrid model architecture. Although some of these issues are problem-specific, they again speak to general issues in data-driven modelling concerning the data available for training, especially in infrequently observed conditions, sparse sampling, and the selection of variables to include in the model.

A concern surrounding DL models is that they may identify patterns in a way that is not constrained by known physical laws (what Reichstein and others 2019 call ‘physical inconsistency’). Karniadakis and others (2021) describe three ways that information can be introduced to machine learning (including DL) models to make them ‘physics-informed’: (i) observational biases where the data used to train the model carry information about the underlying mechanisms, (ii) inductive biases where known physical laws are embedded in the model architecture, and (iii) learning biases where the model is penalised for violating physical constraints. Arguably, using inductive bias is the approach that will most strictly honour physical reality but it requires a rather complete mechanistic understanding of the system (difficult for complex and open systems such as ecosystems) and does not scale well (Karniadakis and others 2021). These physics-informed methods are beginning to be adopted by ecosystem ecologists although the terminology used differs between disciplines and applications. For example, building on the ‘theory-guided data science’ of Karpatne and others (2017), Jia and others (2019) implement a RNN to predict lake water temperature in a way that honours the

conservation of energy and the relationships between depth and density. Their constrained RNN outperformed a physical lake ecosystem model, and Jia and others (2019) argue that the inclusion of physical constraints makes it more easily generalisable. Read and others (2019) tested this approach across more lakes, comparing it to an unsupervised DL model and a physical model of lake temperature. Their hybrid DL model outperformed the others both in lakes where there was detailed site-specific information and in a wider pool of nearly 70 test lakes where there was less information. Likewise, physical laws might also be used to evaluate model performance; for example, Razavi (2021) show how a DL model of precipitation could be tested using a temperature threshold for snow formation (that is, is snow vs. rain predicted at appropriate temperatures). Ultimately, linking DL and mechanistic models may improve predictive performance and help develop causal understanding of the systems of interest.

### Meta-models and Model Emulation

Another potential application of DL in models of ecosystem dynamics is as model emulators or meta-models. Even with access to large-scale computing infrastructure, there are limits to which brute-force approaches can run complex ecological models over large areas and/ or long periods. Many techniques have been proposed for scaling models before, during, or after model application (Fritsch and others 2020), including meta-modelling (Urban and others 1999; Cipriotti and others 2015) or model emulation (Reichstein and others 2019). The basis of meta-modelling is that a simpler (in computational or representational terms) form of a complex model is developed and applied over larger, longer, or more heterogeneous conditions, or used in what would be otherwise unfeasible computational experiments. For example, Cipriotti and others (2015) used matrix models to synthesise a complex individual-based model of grassland dynamics by tracking transitions between states in grid cells. DL models provide a way to deal with cases with many states and a more complex environment, in which full coverage of all possible combinations is impossible by conventional approaches. Rammer and Seidl (2019a) used a DNN that learns the probability of transitions between  $10^3$  and  $10^6$  ecosystem states from process-based simulations conditional on state history, spatial context, and environmental conditions. The approach was subsequently used to project post-fire regeneration under future climate and fire regimes

for the Greater Yellowstone Ecosystem (USA), projecting substantial regeneration failure in the twenty-first century due to limited seed supply and post-fire drought (Rammer and others 2021). Similarly, Dagon and others (2020) trained a feed-forward neural network to emulate a detailed model of ecosystem fluxes at extended spatial scales.

### Challenges for DL in Ecosystem Ecology

Deep learning has considerable potential for ecosystem ecology, as illustrated for the three application domains described above. However, considerable challenges remain. Here we consider three challenges for the use of DL in ecosystem ecology and discuss potential ways to mitigate them: (i) data availability, and especially large labelled databases for supervised learning, (ii) the issues of interpretability in data-led modelling (for example, understanding *why* a model makes a given prediction), and (iii) the environmental costs of data-led methods.

### Dealing with a Paucity of High-Quality (Labelled) Data

Most applications of DL by ecosystem ecologists have involved supervised classification; in other words, a model learns its task using a labelled or annotated training (reference) dataset. However, supervised learning depends on the availability and veracity of large labelled datasets (Karpatne and others 2019), especially given the concern that DL models may overfit when trained on small datasets (Goodfellow and others 2016). In a number of examples reviewed earlier, model performance was negatively affected by scarce and patchy data, especially for extreme conditions. There is a massive effort involved in developing expert-curated training sets, whether ecological or not. Citizen science may provide one solution; the Serengeti snapshot database (Swanson and others 2015), for example, contains 3.2 million images of animals across 1.2 million snapshot captures, which have been labelled (presence, identification, count) by volunteers at an estimated cost of 14.6 years' worth of 40-h weeks. However, while citizen science initiatives may increase the scope of such efforts they will also potentially carry biases in space, time and expertise, although this will vary with the project and may not differ from 'professional' data (Kosmala and others 2016). The effort to measure plant functional traits by integrating data from the open TRY database and the citizen science application



iNaturalist described by Schiller and others (2021) is an interesting example of how different data streams can be used to develop global syntheses. Irrespective of such efforts, there are many ecological contexts where there will be a persistent shortage of high-quality labelled.

Various solutions have been proposed to address the issue of limited training data based on concerns that models trained on small datasets are vulnerable to overfitting. First, although the large majority of ecological applications use supervised learning, the development of unsupervised and self-supervised algorithms that circumvent the need for extensive labelled training data is an active area of research (for example, Yan and Wang 2022). Where supervised models are used, two solutions to data paucity are generating synthetic data to augment existing databases and minimising the amount of labelled data required. Data augmentation is the generation of new training data from existing training examples. For example, images can be geometrically altered (shifting, mirroring, rotating, zooming, shearing) or audio data distorted to increase the data set while not having to add more raw information or labelling effort. This approach has received some attention from ecosystem ecologists. For example, Grünig and others (2021) used data augmentation to expand the data available to train a model for detecting and classifying damage to plants by pests and pathogens. Another alternative, especially for temporal data is to use the output of physical simulations to train DL models (a form of meta-modelling); of course, using a process-based model to train a DL relies on the robustness and/or the transferability of the physical model.

Another way to deal with the problem of the data required to train effective DL models is to limit the amount of labelling required. Two approaches that seek to achieve this are transfer learning and active learning. Transfer learning takes advantage of models developed for one specific setting elsewhere (Goodfellow and others 2016; Weiss and others 2016). Transfer learning has three potential benefits compared to training a new model 'from scratch' (Torrey and Shavlik 2010): better initial performance, more rapid improvement in performance as the model is trained, and better final performance. Transfer learning leverages the property that in broad problem domains (for example, image classification) the early layers are often similar across DL models, irrespective of the specific problem (Yosinski and others 2014). By using pre-trained models as the starting point for model training, knowledge can be transferred (for

example, general image understanding in the context of a DL model) to a new task where there is limited labelled data (Goodfellow and others 2016). Another approach aiming to reduce the labelling burden is active learning, that uses methods to select the most informative examples (that is, those from which the DNN can learn the most at a given point in time) from the pool of unlabelled data. In an iterative process an expert user is occasionally asked to label such informative unclassified samples during model training (Norouzzadeh and others 2021), thus selectively extending the data set. The hope is that by being selective about which data are labelled by the expert (the so-called oracle) the costs involved will drop as a reduced set of the most informative data are selected for annotation.

Ecologists have already begun to use active and transfer learning. For example, Valan and others (2019) used transfer learning in the taxonomic identification of invertebrates (via a CNN) because of a lack of training data and concern over the computational cost of fine-tuning. They used a pre-trained CNN trained on the ImageNet data set (currently  $1.4 \times 10^7$  images in 100,000 classes), extracted features (that is, an intermediate representation in the CNN) and used these to train a support vector machine with a smaller labelled dataset (100–1000 s of images). Transfer learning can involve models trained on quite different data. Norouzzadeh and others (2018) tested the ability of their models of wildlife imagery when trained on smaller datasets simulating wildlife cameras and the generic ImageNet database, which is not wildlife-specific. In both cases, the models performed well. Russo and others (2020) tested the effectiveness of active learning to reduce the labelling effort involved in detecting anomalies in data (in their case, specific conductivity in mesocosm experiments). Their workflow involved labelling data (complete labelling, random labelling of a subset, active learning) and then training DNN models using these labels. Their analysis demonstrates that models with high predictive accuracy can be developed with a fraction of the labelling effort using an active learning method. Likewise, Norouzzadeh and others (2021) demonstrate how a workflow integrating active learning can massively reduce labelling requirements; the most accurate of the algorithm they used had an accuracy comparable to that of Norouzzadeh and others (2018) but labelling just 14,000 versus 3.2 million images (a 99% reduction).

## Prediction, Explanation, Interpretability, and Learning

Ecological modellers have long debated the relative merits of simple and complex models in various guises such as the realism versus tractability trade-off (Levins 1966; Evans and others 2013; Razavi 2021). This argument is particularly acute for deep learning methods, especially given their seemingly “unreasonable effectiveness” (Sejnowski 2020) and the large amounts of data they typically require. While in some problem domains explainability may not matter, in others it does. Thus, there is growing interest in ‘interpretable machine learning’ (Murdoch and others 2019). Roscher and others (2020) distinguish between: (i) transparency (being able to communicate the decisions made in the model implementation process and how they influence the outcomes), (ii) interpretability (for example, using post hoc assessment to understand how a decision based on a model prediction was reached) and (iii) explainability (explaining the outcome of a modelling exercise in a process-sense, acknowledging the context-dependent nature of explanation). Methods designed to help a modelling exercise develop these qualities have begun to be used by ecologists. These methods can examine the global (how the model learned to identify patterns from the data it was trained with) or local model structure (why did a model make a prediction for a given site or sample), and are reviewed in detail in an ecological context by Lucas (2020). Ryo and others (2021) illustrate the use of these interpretation methods (‘explainable AI’) in the context of species distribution models and highlight how explaining the global model and individual predictions can yield improved causal understanding of the system being predicted. Other examples include the visual interrogation of DL models using saliency maps, which depict how each data point influences the nodes in a DNN, or by methods that highlight surprising predictions (McGovern and others 2019). Likewise, sensitivity analysis and layerwise relevance propagation can facilitate understanding of a model’s outcomes by mapping the relationship between inputs and outputs (Montavon and others 2018; Toms and others 2020). These methods and others more routinely applied to machine learning approaches (for example, variable importance metrics or partial dependence plots) help understand the model (interpretability) but do not necessarily generate knowledge of themselves. Thus, as Roscher and others (2020) and Razavi (2021) emphasise, domain-specific expertise remains crucial for inter-

preting and assessing DL models’ credibility and predictions.

The ability of DL to uncover patterns in large, messy and heterogeneous data may inspire new hypotheses that can be tested with experiments or models. Identifying surprising predictions (or ones that have not been observed) is important because one route to model-based learning is for such predictions to be empirically confirmed (Mankin and others 1975). As Reichstein and others (2019) outline, this does not necessarily challenge the ‘classical’ hypothetico-deductive model; instead, the patterns identified by DL approaches constitute new ways to observe complex systems. Patterns that cannot be explained by existing theoretical frameworks can guide and inform new experiments. In this way, resource-intensive experiments could be more efficiently targeted. For example, the RNN developed by Kraft and others (2019) to explore how memory effects vary across biomes generates hypotheses about the causes of that biome-level variation (in their case, they speculate that the temporal grain of climate variability will influence the importance of memory effects). Where observations do not depart from existing theory or understanding, they may improve model predictions and parameterisation. In this context, there is likely an important role for unsupervised methods; that is, those where a model makes a prediction to unlabelled data. For example, Sonnewald and others (2020) used machine learning techniques to identify marine eco-provinces from high-dimensional nutrient and plankton data; their approach identified approximately 100 unique eco-provinces. Then, the question for ecosystem ecology is, which biogeochemical and ecological processes and variables control those eco-provinces? Why do they vary through time and space? And how do they relate to existing classifications derived in other ways (for example, via expert assessment)?

## Reconciling Energy and Environmental Costs of Data-Led Approaches

The modelling most frequently conducted by ecologists is not as energy expensive as the approaches used in other fields, such as large-scale natural language models. Yet, in a review of the ecological applications of DL it would be remiss not to touch on recent concerns about the environmental (mainly C and energy) costs of computer-intensive methods (Dhar 2020; Schwartz and others 2020). These concerns are true of any computationally expensive approach, although the size



and training effort in some DL models makes it acute (Thompson and others 2020). The emphasis in DL models has been on extracting maximal predictive performance, which will result in high energy usage. However, as Canziani and others (2017) demonstrate, energy limits probably set upper bounds on practical accuracy given the relationship between time and performance is hyperbolic, and so it becomes necessary to trade off predictive performance and energy cost. Less computationally intensive models may also be more practical for deployment in edge computing (Tuia and others 2022). Recently, guidelines for environmentally sustainable computing have been published (Lannelongue and others 2021) alongside calls for the energy costs of computationally intensive projects to be reported (Lottick and others 2019; Strubell and others 2020). To support efficiency in reporting open-source code bases and online apps (for example, Green Algorithms [<https://green-algorithms.org/>] and machine learning CO<sub>2</sub> Impact [<https://mlco2.github.io/impact/>]) have been developed. We anticipate rapid developments in this area and a move towards an energy conscious 'green AI' (Schwartz and others 2020).

## CONCLUSION

Whether some of the more hyperbolic claims regarding DL will prove well-founded remains to be seen, but there is no doubt that deep learning algorithms offer ecologists opportunities for prediction and understanding in the dawning age of big data. These opportunities range from incremental advances in existing questions (that is, application of DL methods to existing problems), to the expansion of the scope and scale we ask questions at, to entirely new (and unpredictable) questions and processing capacities. We anticipate that hybrid physical-DL models are an area where there are particular opportunities for ecosystem ecology as the binary view of mechanistic vs. empirical models becomes blurred. However, DL methods also amplify debates about the place of data, theory and models in science. To understand data, do we need a hypothetical generating model? Or can we identify empirical truisms to make predictions? These questions are long-standing and likely unresolvable; focusing on them might be unhelpful for advancing the science of ecology, particularly if they are posed as binaries. Thus, the challenge for ecosystem ecologists in leveraging data-led approaches is not solely technical but is also to reconcile competing narratives in ways that

equip us to deal with a rapidly changing environment.

## ACKNOWLEDGEMENTS

R.S. and W.R. acknowledge support from the European Research Council (ERC) under the European Union's Horizon 2020 research and innovation programme (Grant Agreement No. 101001905). We thank three anonymous referees for their insightful comments that helped us to improve the manuscript.

## FUNDING

Open Access funding enabled and organized by CAUL and its Member Institutions.

*Open Access* This article is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons licence, and indicate if changes were made. The images or other third party material in this article are included in the article's Creative Commons licence, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons licence and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this licence, visit <http://creativecommons.org/licenses/by/4.0/>.

## REFERENCES

- Baldocchi DD. 2020. How eddy covariance flux measurements have contributed to our understanding of Global Change Biology. *Global Change Biology* 26:242–260. <https://doi.org/10.1111/gcb.14807>. Last accessed 23/05/2022.
- Besnard S, Carvalhais N, Arain MA, Black A, Brede B, Buchmann N, Chen J, Clevers JGPW, Dutrieux LP, Gans F, Herold M, Jung M, Kosugi Y, Knohl A, Law BE, Paul-Limoges E, Lohila A, Merbold L, Roupsard O, Valentini R, Wolf S, Zhang X, Reichstein M. 2019. Memory effects of climate and vegetation affecting net ecosystem CO<sub>2</sub> fluxes in global forests. *PLoS ONE* 14:e0211510. <https://doi.org/10.1371/journal.pone.0211510>. Last accessed 01/10/2021.
- Borowiec ML, Dikow RB, Frandsen PB, McKeeken A, Valentini G, White AE. 2022. Deep learning as a tool for ecology and evolution. *Methods in Ecology and Evolution* 13:1640–60. <https://doi.org/10.1111/2041-210X.13901>. Last accessed 25/08/2022.

- Brodrick PG, Davies AB, Asner GP. 2019. Uncovering ecological patterns with convolutional neural networks. *Trends in Ecology & Evolution* 34:734–45. <https://linkinghub.elsevier.com/retrieve/pii/S0169534719300862>. Last accessed 29/09/2021
- Canziani A, Paszke A, Culurciello E. 2017. An analysis of deep neural network models for practical applications. <http://arxiv.org/abs/1605.07678>. Last accessed 23/05/2022
- Chen H, Wang Y, Guo T, Xu C, Deng Y, Liu Z, Ma S, Xu C, Xu C, Gao W. 2021a. Pre-trained image processing transformer. In: IEEE/CVF conference on computer vision and pattern recognition (CVPR). Los Alamitos, CA, USA: IEEE Computer Society. pp 12294–305. <https://doi.org/10.1109/CVPR46437.2021.01212>
- Chen J, Dafflon B, Tran AP, Falco N, Hubbard SS. 2021b. A deep learning hybrid predictive modeling (HPM) approach for estimating evapotranspiration and ecosystem respiration. *Hydrology and Earth System Sciences* 25:6041–66. <https://hess.copernicus.org/articles/25/6041/2021b/>. Last accessed 23/05/2022
- Christin S, Hervet É, Lecomte N. 2019. Applications for deep learning in ecology. *Methods in Ecology and Evolution* 10:1632–1644. <https://doi.org/10.1111/2041-210X.13256?af=R>
- Cipriotti PA, Wiegand T, Pütz S, Bartoloni NJ, Paruelo JM. 2015. Nonparametric upscaling of stochastic simulation models using transition matrices. *Methods Ecol Evol* 7:313–322. <https://doi.org/10.1111/2041-210X.12464>
- Cleverly J, Eamus DJ, Edwards W, Grant M, Grundy MJ, Held A, Karan M, Lowe AJ, Prober SM, Sparrow B, Morris B. 2019. TERN, Australia's land observatory: addressing the global challenge of forecasting ecosystem responses to climate variability and change. *Environ Res Lett* 14:095004. <https://doi.org/10.1088/1748-9326/ab33cb>. Last accessed 24/08/2022
- Dagon K, Sanderson BM, Fisher RA, Lawrence DM. 2020. A machine learning approach to emulation and biophysical parameter estimation with the Community Land Model, version 5. *Advances in Statistical Climatology, Meteorology and Oceanography* 6:223–44. <https://ascmo.copernicus.org/articles/6/223/2020/>. Last accessed 23/05/2022
- Daood A, Ribeiro E, Bush M. 2016. Pollen grain recognition using deep learning. In: Bebis G, Boyle R, Parvin B, Koracin D, Porikli F, Skaff S, Entezari A, Min J, Iwai D, Sadagic A, Scheidegger C, Isenberger T, editors. *Advances in Visual Computing. Lecture Notes in Computer Science*. Cham: Springer International Publishing. pp 321–30
- Dhar P. 2020. The carbon impact of artificial intelligence. *Nature Machine Intelligence* 2:423–5. <https://www.nature.com/articles/s42256-020-0219-9>. Last accessed 30/09/2021.
- Dietze MC, Fox A, Beck-Johnson LM, Betancourt JL, Hooten MB, Jarnevich CS, Keitt TH, Kenney MA, Laney CM, Larsen LG, Loescher HW, Lunch CK, Pijanowski BC, Randerson JT, Read EK, Tredennick AT, Vargas R, Weathers KC, White EP. 2018. Iterative near-term ecological forecasting: Needs, opportunities, and challenges. *Proc Natl Acad Sci USA* 115:1424–32. <https://doi.org/10.1073/pnas.1710231115>. Last accessed 29/09/2021
- Evans MR, Grimm V, Johst K, Knuuttila T, de Langhe R, Lessells CM, Merz M, O'Malley MA, Orzack SH, Weisberg M, Wilkinson DJ, Wolkenhauer O, Benton TG. 2013. Do simple models lead to generality in ecology? *Trends in Ecology & Evolution* 28:578–83. <http://linkinghub.elsevier.com/retrieve/pii/S0169534713001444>. Last accessed 28/07/2015.
- Farley SS, Dawson A, Goring SJ, Williams JW. 2018. Situating ecology as a big-data science: current advances, challenges, and solutions. *BioScience* 68:563–76. <https://academic.oup.com/bioscience/article/68/8/563/5049569>. Last accessed 29/09/2021.
- Fritsch M, Lischke H, Meyer KM. 2020. Scaling methods in ecological modelling. *Methods in Ecology and Evolution* 11:1368–1378. [https://doi.org/10.1111/2041-210X.13466?af=R&utm\\_source=researcher\\_app&utm\\_medium=referral&utm\\_campaign=RESR\\_MRKT\\_Researcher\\_inbound](https://doi.org/10.1111/2041-210X.13466?af=R&utm_source=researcher_app&utm_medium=referral&utm_campaign=RESR_MRKT_Researcher_inbound)
- Goodfellow I, Bengio Y, Courville A. 2016. *Deep Learning*. Cambridge: MIT Press.
- Graving JM, Chae D, Naik H, Li L, Koger B, Costelloe BR, Couzin ID. 2019. DeepPoseKit, a software toolkit for fast and robust animal pose estimation using deep learning. Baldwin IT, Shaevitz JW, Shaevitz JW, Stephens G, editors. *eLife* 8:e47994. <https://doi.org/10.7554/eLife.47994>. Last accessed 24/05/2022.
- Grünig M, Razavi E, Calanca P, Mazzi D, Wegner JD, Pellissier L. 2021. Applying deep neural networks to predict incidence and phenology of plant pests and diseases. *Ecosphere* 12:e03791. <https://doi.org/10.1002/ecs2.3791>. Last accessed 12/05/2022.
- Hansen WD, Schwartz NB, Williams AP, Albrich K, Kueppers LM, Rammig A, Reyer CPO, Staver AC, Seidl R. 2022. Global forests are influenced by the legacies of past inter-annual temperature variability. *Environ Res: Ecology* 1:011001. <https://doi.org/10.1088/2752-664X/ac6e4a>. Last accessed 01/09/2022.
- Irrgang C, Boers N, Sonnewald M, Barnes EA, Kadow C, Staneva J, Saynisch-Wagner J. 2021. Towards neural Earth system modelling by integrating artificial intelligence in Earth system science. *Nat Mach Intell* 3:667–74. <https://www.nature.com/articles/s42256-021-00374-3>. Last accessed 30/09/2021
- Jia X, Willard J, Karpatne A, Read J, Zwart J, Steinbach M, Kumar V. 2019. Physics guided RNNs for modeling dynamical systems: a case study in simulating lake temperature profiles. In: *Proceedings of the 2019 SIAM International Conference on Data Mining (SDM)*. Proceedings. Society for Industrial and Applied Mathematics. pp 558–66. <https://doi.org/10.1137/1.9781611975673.63>. Last accessed 23/05/2022
- Karniadakis GE, Kevrekidis IG, Lu L, Perdikaris P, Wang S, Yang L. 2021. Physics-informed machine learning. *Nat Rev Phys* 3:422–40. <https://www.nature.com/articles/s42254-021-00314-5>. Last accessed 25/08/2022
- Karpatne A, Atluri G, Faghmous JH, Steinbach M, Banerjee A, Ganguly A, Shekhar S, Samatova N, Kumar V. 2017. Theory-guided data science: a new paradigm for scientific discovery from data. *IEEE Transactions on Knowledge and Data Engineering* 29:2318–2331.
- Karpatne A, Ebert-Uphoff I, Ravela S, Babaie HA, Kumar V. 2019. Machine learning for the geosciences: challenges and opportunities. *IEEE Transactions on Knowledge and Data Engineering* 31:1544–1554.
- Kattenborn T, Eichel J, Wiser S, Burrows L, Fassnacht FE, Schmidtlein S. 2020. Convolutional Neural Networks accurately predict cover fractions of plant species and communities in Unmanned Aerial Vehicle imagery. *Remote Sensing in Ecology and Conservation* 6:472–86. <https://doi.org/10.1002/rse2.146>. Last accessed 06/10/2021
- Kattge J, Bönsch G, Díaz S, Lavorel S, Prentice IC, Leadley P, Tautenhahn S, Werner GDA, Aakala T, Abedi M, Acosta ATR, Adamidis GC, Adamson K, Aiba M, Albert CH, Alcántara JM, Alcázar C C, Aleixo I, Ali H, Amiaud B, Ammer C, Amoroso

- MM, Anand M, Anderson C, Anten N, Antos J, Apgaua DMG, Ashman T-L, Asmara DH, Asner GP, Aspinwall M, Atkin O, Aubin I, Bastrup-Spohr L, Bahalkeh K, Bahn M, Baker T, Baker WJ, Bakker JP, Baldocchi D, Baltzer J, Banerjee A, Baranger A, Barlow J, Barneche DR, Baruch Z, Bastianelli D, Battles J, Bauerle W, Bauters M, Bazzato E, Beckmann M, Beeckman H, Beierkuhnlein C, Bekker R, Belfry G, Belluau M, Beloiu M, Benavides R, Benomar L, Berdugo-Lattke ML, Berenguer E, Bergamin R, Bergmann J, Bergmann Carlucci M, Berner L, Bernhardt-Römermann M, Bigler C, Bjorkman AD, Blackman C, Blanco C, Blonder B, Blumenthal D, Bocanegra-González KT, Boeckx P, Bohlman S, Böhning-Gaese K, Boisvert-Marsh L, Bond W, Bond-Lamberty B, Boom A, Boonman CCF, Bordin K, Boughton EH, Boukili V, Bowman DMJS, Bravo S, Brendel MR, Broadley MR, Brown KA, Bruelheide H, Brunnich F, Bruun HH, Bruy D, Buchanan SW, Bucher SF, Buchmann N, Buitenwerf R, and others 2020. TRY plant trait database – enhanced coverage and open access. *Global Change Biology* 26:119–88. <https://doi.org/10.1111/gcb.14904>. Last accessed 24/05/2022.
- Kays R, McShea WJ, Wikelski M. 2020. Born-digital biodiversity data: Millions and billions. *Diversity and Distributions* 26:644–648. <https://doi.org/10.1111/ddi.12993>.
- Keitt TH, Abelson ES. 2021. Ecology in the age of automation. *Science* 373:858–9. <https://doi.org/10.1126/science.abi4692>. Last accessed 11/05/2022.
- Keller M, Schimel DS, Hargrove WW, Hoffman FM. 2008. A continental strategy for the National Ecological Observatory Network. *Frontiers in Ecology and the Environment* 6:282–4. <https://doi.org/10.1890/1540-9295%282008%296%5B282%3AACSTN%5D2.0.CO%3B2>. Last accessed 24/08/2022.
- Koppa A, Rains D, Hulsman P, Poyatos R, Miralles DG. 2022. A deep learning-based hybrid model of global terrestrial evaporation. *Nat Commun* 13:1912. <https://www.nature.com/articles/s41467-022-29543-7>. Last accessed 23/05/2022.
- Kosmala M, Wiggins A, Swanson A, Simmons B. 2016. Assessing data quality in citizen science. *Frontiers in Ecology and the Environment* 14:551–60. <https://doi.org/10.1002/fee.1436>. Last accessed 12/05/2022.
- Kraft B, Jung M, Körner M, Requena Mesa C, Cortés J, Reichstein M. 2019. Identifying dynamic memory effects on vegetation state using recurrent neural networks. *Frontiers in Big Data* 2:31. <https://doi.org/10.3389/fdata.2019.00031/full>. Last accessed 01/10/2021.
- LaDeau SL, Han BA, Rosi-Marshall EJ, Weathers KC. 2017. The next decade of big data in ecosystem science. *Ecosystems* 20:274–83. <https://doi.org/10.1007/s10021-016-0075-y>. Last accessed 28/09/2021.
- Lannelongue L, Grealey J, Bateman A, Inouye M. 2021. Ten simple rules to make your computing more environmentally sustainable. *PLOS Computational Biology* 17:e1009324. <https://journals.plos.org/ploscompbiol/article?id=https://doi.org/10.1371/journal.pcbi.1009324>. Last accessed 07/10/2021.
- LeCun Y, Bengio Y, Hinton G. 2015. Deep learning. *Nature* 521:436–44. <http://www.nature.com/articles/nature14539>. Last accessed 15/09/2021.
- Levins R. 1966. The strategy of model building in population biology. *American Scientist* 54:421–31. <https://mechanism.ucsd.edu/teaching/models/levins.modelbuilding.pdf>.
- Lopez-Marcano S, L. Jinks E, Buelow CA, Brown CJ, Wang D, Kusy B, M. Ditria E, Connolly RM. 2021. Automatic detection of fish and tracking of movement for ecology. *Ecol Evol* 11:8254–63. <https://www.ncbi.nlm.nih.gov/pmc/articles/PMC8216886/>. Last accessed 24/05/2022.
- Lottick K, Susai S, Friedler SA, Wilson JP. 2019. Energy Usage Reports: Environmental awareness as part of algorithmic accountability. arXiv:191108354 [cs, stat]. <http://arxiv.org/abs/1911.08354>. Last accessed 07/10/2021.
- Lucas TCD. 2020. A translucent box: interpretable machine learning in ecology. *Ecol Monogr* 90. <https://doi.org/10.1002/ecm.1422>.
- Lürig MD, Donoughe S, Svensson EI, Porto A, Tsuboi M. 2021. Computer vision, machine learning, and the promise of phenomics in ecology and evolutionary biology. *Frontiers in Ecology and Evolution* 9. <https://doi.org/10.3389/fevo.2021.642774>. Last accessed 11/05/2022.
- Mankin JB, O'Neill RV, Shugart HH, Rust BW. 1975. The importance of validation in ecosystems analysis. In: Innis GS, editor. *New Directions in the Analysis of Ecological Systems, Part 1*. LaJolla, California: Simulation Councils Proceedings Series. Society for Computer Simulation (Simulation Councils). pp 63–71.
- McGovern A, Lagerquist R, Gagne DJ, Jergensen GE, Elmore KL, Homeyer CR, Smith T. 2019. Making the black box more transparent: understanding the physical implications of machine learning. *Bulletin of the American Meteorological Society* 100:2175–99. <https://journals.ametsoc.org/view/journals/bams/100/11/bams-d-18-0195.1.xml>. Last accessed 30/09/2021.
- Montavon G, Samek W, Müller K-R. 2018. Methods for interpreting and understanding deep neural networks. *Digital Signal Processing* 73:1–15. <https://www.sciencedirect.com/science/article/pii/S1051200417302385>. Last accessed 30/09/2021.
- Murdoch WJ, Singh C, Kumbier K, Abbasi-Asl R, Yu B. 2019. Definitions, methods, and applications in interpretable machine learning. *Proc Natl Acad Sci USA* 116:22071–80. <https://doi.org/10.1073/pnas.1900654116>. Last accessed 29/09/2021.
- Norouzzadeh MS, Morris D, Beery S, Joshi N, Jojic N, Clune J. 2021. A deep active learning system for species identification and counting in camera trap images. *Methods in Ecology and Evolution* 12:150–61. <https://doi.org/10.1111/2041-210X.13504>. Last accessed 06/10/2021.
- Norouzzadeh MS, Nguyen A, Kosmala M, Swanson A, Palmer MS, Packer C, Clune J. 2018. Automatically identifying, counting, and describing wild animals in camera-trap images with deep learning. *Proceedings of the National Academy of Sciences* 115:E5716–25. <https://doi.org/10.1073/pnas.1719367115>. Last accessed 29/09/2021.
- Olden JD, Lawler JJ, Poff NL. 2008. Machine learning methods without tears: a primer for ecologists. *The Quarterly Review of Biology* 83:171–193. <https://doi.org/10.1086/587826>.
- Olsson O, Karlsson M, Persson AS, Smith HG, Varadarajan V, Yourstone J, Stjernman M. 2021. Efficient, automated and robust pollen analysis using deep learning. *Methods in Ecology and Evolution* 12:850–62. <https://doi.org/10.1111/2041-210X.13575>. Last accessed 06/10/2021.
- Rahmani F, Lawson K, Ouyang W, Appling A, Oliver S, Shen C. 2021. Exploring the exceptional performance of a deep learning stream temperature model and the value of stream-flow data. *Environ Res Lett* 16:024025. <https://doi.org/10.1088/1748-9326/abd501>.
- Rammer W, Braziunas KH, Hansen WD, Ratajczak Z, Westerling AL, Turner MG, Seidl R. 2021. Widespread regeneration fail-



- ure in forests of Greater Yellowstone under scenarios of future climate and fire. *Glob Change Biol* 27:4339–51. <https://doi.org/10.1111/gcb.15726>. Last accessed 27/09/2021
- Rammer W, Seidl R. 2019a. A scalable model of vegetation transitions using deep neural networks. *Methods in Ecology and Evolution* 10:879–890. <https://doi.org/10.1111/2041-210X.13171?af=R>.
- Rammer W, Seidl R. 2019b. Harnessing deep learning in ecology: an example predicting bark beetle outbreaks. *Frontiers in Plant Science* 10:1327. <https://doi.org/10.3389/fpls.2019.01327>. Last accessed 06/10/2021
- Rawat W, Wang Z. 2017. Deep convolutional neural networks for image classification: a comprehensive review. *Neural Computation* 29:2352–2449.
- Razavi S. 2021. Deep learning, explained: Fundamentals, explainability, and bridgeability to process-based modelling. *Environmental Modelling & Software* 144:105159. <https://www.sciencedirect.com/science/article/pii/S1364815221002024>. Last accessed 24/08/2022
- Read JS, Jia X, Willard J, Appling AP, Zwart JA, Oliver SK, Karpatne A, Hansen GJA, Hanson PC, Watkins W, Steinbach M, Kumar V. 2019. Process-guided deep learning predictions of lake water temperature. *Water Resources Research* 55:9173–90. <https://doi.org/10.1029/2019WR024922>. Last accessed 23/05/2022.
- Reichstein M, Camps-Valls G, Stevens B, Jung M, Denzler J, Carvalhais N, Prabhat. 2019. Deep learning and process understanding for data-driven Earth system science. *Nature* 566:195–204. <http://www.nature.com/articles/s41586-019-0912-1>. Last accessed 16/12/2020.
- Roscher R, Bohn B, Duarte MF, Garcke J. 2020. Explainable machine learning for scientific insights and discoveries. *IEEE Access* 8:42200–42216.
- Russo S, Lürig M, Hao W, Matthews B, Villez K. 2020. Active learning for anomaly detection in environmental data. *Environmental Modelling & Software* 134:104869. <https://www.sciencedirect.com/science/article/pii/S1364815220309269>. Last accessed 23/05/2022
- Ryo M, Angelov B, Mammola S, Kass JM, Benito BM, Hartig F. 2021. Explainable artificial intelligence enhances the ecological interpretability of black-box species distribution models. *Ecography* 44:199–205. <https://doi.org/10.1111/ecog.05360>. Last accessed 01/11/2021.
- Schiller C, Schmidlein S, Boonman C, Moreno-Martínez A, Kattenborn T. 2021. Deep learning and citizen science enable automated plant trait predictions from photographs. *Sci Rep* 11:16395. <https://www.nature.com/articles/s41598-021-95616-0>. Last accessed 24/05/2022
- Schwartz R, Dodge J, Smith NA, Etzioni O. 2020. Green AI. *Communications of the ACM* 63:54–63. <https://doi.org/10.1145/3381831>.
- Sejnowski TJ. 2020. The unreasonable effectiveness of deep learning in artificial intelligence. *Proceedings of the National Academy of Sciences* 117:30033–8. <https://www.pnas.org/doi/https://doi.org/10.1073/pnas.1907373117>. Last accessed 12/05/2022.
- Sonnefeld M, Dutkiewicz S, Hill C, Forget G. 2020. Elucidating ecological complexity: Unsupervised learning determines global marine eco-provinces. *Science Advances* 6:eaay4740. <https://doi.org/10.1126/sciadv.aay4740>. Last accessed 30/09/2021.
- Strubell E, Ganesh A, McCallum A. 2020. Energy and policy considerations for modern deep learning research. *Proceedings of the AAAI Conference on Artificial Intelligence* 34:13693–6. <https://ojs.aaai.org/index.php/AAAI/article/view/7123>. Last accessed 30/09/2021.
- Swanson A, Kosmala M, Lintott C, Simpson R, Smith A, Packer C. 2015. Snapshot Serengeti, high-frequency annotated camera trap images of 40 mammalian species in an African savanna. *Sci Data* 2:1–14. <http://www.nature.com/articles/sdata201526>. Last accessed 30/09/2021.
- Thompson NC, Greenewald K, Lee K, Manso GF. 2020. The computational limits of deep learning. arXiv:200705558 [cs, stat]. <http://arxiv.org/abs/2007.05558>. Last accessed 01/11/2021..
- Toms BA, Barnes EA, Ebert-Uphoff I. 2020. Physically interpretable neural networks for the geosciences: applications to earth system variability. *Journal of Advances in Modeling Earth Systems* 12:e2019MS002002. <https://doi.org/10.1029/2019MS002002>. Last accessed 07/10/2021.
- Torrey L, Shavlik J. 2010. Transfer learning. In: *Handbook of Research on Machine Learning Applications and Trends: Algorithms, Methods, and Techniques*. IGI global. pp 242–64.
- Tuia D, Kellenberger B, Beery S, Costelloe BR, Zuffi S, Risse B, Mathis A, Mathis MW, van Langevelde F, Burghardt T, Kays R, Klinck H, Wikelski M, Couzin ID, van Horn G, Crofoot MC, Stewart CV, Berger-Wolf T. 2022. Perspectives in machine learning for wildlife conservation. *Nat Commun* 13:792. <http://www.nature.com/articles/s41467-022-27980-y>. Last accessed 11/05/2022
- Urban DL, Acevedo MF, Garman SL. 1999. Scaling fine-scale processes to large-scale patterns using models derived from models: meta-models. *Spatial Modeling of Forest Landscapes: Approaches and Applications*. Cambridge University Press: Cambridge. pp 125–163.
- Valan M, Makonyi K, Maki A, Vondráček D, Ronquist F. 2019. Automated taxonomic identification of insects with expert-level accuracy using effective feature transfer from convolutional networks. *Systematic Biology* 68:876–95. <https://doi.org/10.1093/sysbio/syz014>. Last accessed 17/11/2021.
- Vaswani A, Shazeer N, Parmar N, Uszkoreit J, Jones L, Gomez AN, Kaiser Ł, Polosukhin I. 2017. Attention is all you need. In: *Proceedings of the 31st International Conference on Neural Information Processing Systems*. NIPS'17. Red Hook, NY, USA: Curran Associates Inc. pp 6000–10.
- Walter T, Couzin ID. 2021. TRex, a fast multi-animal tracking system with markerless identification, and 2D estimation of posture and visual fields. Lentinck D, Rutz C, Pujades S, editors. *eLife* 10:e64000. <https://doi.org/10.7554/eLife.64000>. Last accessed 24/05/2022
- Wang D, Barabási A-L. 2021. *The Science of Science*, 1st edn. Cambridge: Cambridge University Press.
- Weeks BC, Zhou Z, O'Brien BK, Darling R, Dean M, Dias T, Hassena G, Zhang M, Fouhey DF. 2022. A deep neural network for high-throughput measurement of functional traits on museum skeletal specimens. *Methods in Ecology and Evolution* n/a. <https://doi.org/10.1111/2041-210X.13864>. Last accessed 24/05/2022
- Weiss K, Khoshgoftaar TM, Wang D. 2016. A survey of transfer learning. *Journal of Big Data* 3:9. <https://doi.org/10.1186/s40537-016-0043-6>. Last accessed 23/05/2022.
- Xu T, Longyang Q, Tyson C, Zeng R, Neilson BT. 2022. Hybrid physically based and deep learning modeling of a snow dominated, mountainous, karst watershed. *Water Resources Research* 58:e2021WR030993. <https://doi.org/10.1029/2021WR030993>. Last accessed 23/05/2022.

Yan J, Wang X. 2022. Unsupervised and semi-supervised learning: the next frontier in machine learning for plant systems biology. *The Plant Journal* n/a. <https://doi.org/10.1111/tpj.15905>. Last accessed 30/08/2022.

Yosinski J, Clune J, Bengio Y, Lipson H. 2014. How transferable are features in deep neural networks? In: *Proceedings of the 27th International Conference on Neural Information Pro-*

*cessing Systems - Volume 2. NIPS'14. Cambridge, MA, USA: MIT Press. pp 3320–8.*

Zhi W, Feng D, Tsai W-P, Sterle G, Harpold A, Shen C, Li L. 2021. From hydrometeorology to river water quality: can a deep learning model predict dissolved oxygen at the continental scale? *Environmental Science & Technology* 55:2357–2368. <https://doi.org/10.1021/acs.est.0c06783>.