**METHOD**

**IUCNN – Deep learning approaches to approximate species' extinction risk**

**Alexander Zizka**\(^1\,\^2\) | **Tobias Andermann**\(^3\,\^4\) | **Daniele Silvestro**\(^5\,\^6\)

\(^1\)German Center for Integrative Biodiversity Research Halle-Jena-Leipzig (iDiv), University of Leipzig, Leipzig, Germany
\(^2\)Department of Biology, Philipps-University Marburg, Marburg, Germany
\(^3\)Department of Biological and Environmental Sciences, University of Gothenburg, Göteborg, Sweden
\(^4\)Gothenburg Global Biodiversity Centre, Göteborg, Sweden
\(^5\)Department of Biology, University of Fribourg, Fribourg, Switzerland
\(^6\)Swiss Institute of Bioinformatics, Lausanne, Switzerland

**Abstract**

**Aim:** The Red List (RL) from the International Union for the Conservation of Nature is the most comprehensive global quantification of extinction risk, and widely used in applied conservation as well as in biogeographic and ecological research. Yet, due to the time-consuming assessment process, the RL is biased taxonomically and geographically, which limits its application on large scales, in particular for underdocumented areas such as the tropics, or understudied taxa, such as most plants and invertebrates. Here, we present IUCNN, an R-package implementing deep learning models to predict species RL status from publicly available geographic occurrence records (and other data if available).

**Innovation:** We implement a user-friendly workflow to train and validate neural network models, and use them to predict species RL status. IUCNN contains specific functions for extinction risk prediction in the RL framework, including a regression-based approach to account for the ordinal nature of RL categories, a Bayesian approach for improved uncertainty quantification and a convolutional neural network to predict species RL status based on their raw geographic occurrences. Most analyses run with few lines of code, not requiring users to have prior experience with neural network models. We demonstrate the use of IUCNN on an empirical dataset of ~14,000 orchid species, for which IUCNN models can predict extinction risk within minutes, while outperforming comparable methods based on species occurrence information.

**Main conclusions:** IUCNN harnesses innovative methodology to estimate the RL status of large numbers of species. By providing estimates of the number and identity of threatened species in custom geographic or taxonomic datasets, IUCNN enables large-scale automated assessments of the extinction risk of species so far not well represented on the official RL.

**KEYWORDS**
automated assessment, conservation assessment, IUCN, machine learning, red list, R-package

This is an open access article under the terms of the Creative Commons Attribution License, which permits use, distribution and reproduction in any medium, provided the original work is properly cited.
© 2021 The Authors. Diversity and Distributions published by John Wiley & Sons Ltd.
1 | INTRODUCTION

In face of the global biodiversity crisis, the disciplines of biogeography and (macro)ecology can provide an urgently needed perspective for global conservation (Santini et al., 2021; Zizka et al., 2021). Promising contributions include the prioritization of species and areas for conservation and a mechanistic understanding of species extinction risk (Pollock et al., 2020; Rapacciuolo, 2019). Indeed, studies increasingly explore these topics, yet often limited to well-studied regions or taxa, because available information on species extinction risk is scarce and geographically and taxonomically biased (Bachman et al., 2019; Donaldson et al., 2016).

The Red List of the International Union for the Conservation of Nature (RL, www.iucnredlist.org) is arguably the most influential scheme quantifying species extinction risk (Bets et al., 2020) and a prime example for the inclusion of biogeographic principles into conservation. The RL classifies species into five extinction risk categories: least concern (LC), near threatened (NT), vulnerable (VU), endangered (EN), and critically endangered (CR). Professional assessors or specialist groups usually comprised of volunteer scientists classify species into these categories based on at least one of five standardized criteria (“red-listing”). These criteria are based on declines in species populations (Criteria A and C), limited geographic range size (B), small population size (D) or a quantitative extinction risk assessment (E, IUCN, 2012; IUCN Standards & Petitions Subcommittee, 2017). Species that cannot be classified into any of these categories because of a lack of information are considered data deficient (DD); species never considered in a red-listing process are termed not evaluated (NE).

Although the RL is designed for applied conservation, the standardized assessment across regions and taxa makes it a widely used resource for biogeographic and (macro)ecological research. For instance, RL extinction risk assessments have been used to relate traits, such as body mass, to species extinction risk (Boehm et al., 2016; Pincheira-Donoso et al., 2021; Richards et al., 2021; Rolland & Salamin, 2016), to quantify the effect of threats, such as agriculture, on species extinction risk (Polaina et al., 2018), to quantify links between species extinction risk and invasive species (Tingley et al., 2016; Walsh et al., 2012), to characterize the distribution of threatened species (Coll et al., 2015), to predict future biodiversity losses (Andermann et al., 2021; Monroe et al., 2019) and to understand the potential effects of extinction on large-scale diversity patterns (Oliveira et al., 2020; Smiley et al., 2020).

While the above-mentioned examples illustrate the potential of the RL for biogeographic and ecological research, the taxonomic and geographic biases of the RL limit research to well-studied taxa or regions. Red listing is time consuming due to the standardized process and the data requirements which is why only a fraction of the global biodiversity has been evaluated with varying coverage across taxa. For instance, most known vertebrate species (68%) have been evaluated at least once, but the proportion of plants (7%), invertebrates (2%) and fungi & protists (<1%) is lower (Bachman et al., 2019; IUCN, 2018; Lughadha et al., 2020). Furthermore, the proportion of evaluated species is higher in regions with experts and funding available (Bachman et al., 2019), and many of the existing assessments will soon be older than 10 years and thereby, outdated (Rondinini et al., 2014).

To speed up red listing and to overcome these biases, a variety of methods have been developed in recent years to automate the red-listing process. These automation methods trade case-by-case evaluation for reproducibility, scope and speed and they may process thousands of DD or NE species based on publicly available data within minutes (Zizka et al., 2021). There are different flavours of automation methods, which differ in scope, underlying algorithms and data requirements. The most general approaches infer indices required during the formal red-listing procedure, such as the Extent of Occurrence or the Area of Occupancy to support RL assessors (e.g. Bachman et al., 2011), provide preliminary RL assessments based on readily available data following IUCN criteria (e.g. Dauby et al., 2017) or predict species RL category based on species traits (e.g. González-del-Piélgo et al., 2019; Pelletier et al., 2018).

While all of these automation approaches have important limitations (Lughadha et al., 2019; Rivers et al., 2011; Walker et al., 2020), they constitute useful tools for filling gaps in datasets containing DD and NE species. Predictive approaches are particularly promising, since they are able to integrate different data types. For instance, predictions may be based on species geographic distribution, morphology and physiology as well as human disturbance, human use and molecular data (Pelletier et al., 2018; Zizka, Silvestro, et al., 2021). Additionally, predictive approaches can benefit from the active development and ever-improving performance of novel machine learning methods. Among them, neural networks are a highly flexible family of models used to perform classification tasks or regressions (LeCun et al., 2015). Deep neural networks are able to approximate virtually any function and thus provide one of the most general and powerful frameworks for predictions (Goodfellow et al., 2016). While some of the automated assessment methods are implemented in accessible software, most are not. To our knowledge, no application that uses neural networks for RL prediction exists. More generally, the few existing attempts to use machine learning for red listing are documented in scripts from the Supplementary material of research studies, difficult to access for the broader community.

Starting from a deep learning method which we recently applied to predict the RL status of orchid species globally (Zizka, Silvestro, et al., 2021), we here present IUCNN, an R-package implementing neural networks to predict species RL status in an accessible, user-friendly and reproducible way. IUCNN implements a range of models based on fully connected neural networks (NN hereafter) and convolutional neural networks (CNN).

2 | METHODS

A typical workflow in IUCNN contains three steps: 1) label and feature preparation, 2) model training and validation, and 3) prediction.
and visualization (Figure 1), which can be run with few lines of code (Figure 2).

2.1 | Label and feature preparation

\textit{IUCNN} models use features derived from species traits and are trained on species already assessed in the RL. Features may derive from any type of trait but in the simplest case publicly available species occurrence records suffice. Thus, the only input data needed are a set of species with existing RL assessments (the training labels), geo-referenced occurrence records for these species (from which the training features are extracted) and geo-referenced occurrence records for the species whose RL status is to be predicted. In the optimal case, models will be trained on existing RL assessments from evolutionarily related or ecologically similar species (e.g. all members of the orchid family worldwide, to predict extinction risk of orchid species) or related species in the same geographic region (e.g. vascular plants in South America, to predict the extinction risk of South American orchids).

The feature structure and the workflow of feature generation differ between the CNN and NN algorithms. The CNN model implemented in \textit{IUCNN} predicts the RL status of species based on a two-dimensional grid, counting the number of species occurrence records per cell. The \texttt{iucnn\_cnn\_features} function prepares the required input format based on a table of species geo-referenced occurrences and a user-defined coordinate resolution (latitude/longitude) or a user-provided reference raster in any coordinate reference system.

In contrast, NN models require a single vector of features for each species. This means that NN models in \textit{IUCNN} can be trained

\begin{figure}[h]
\centering
\includegraphics[width=\textwidth]{figure1.png}
\caption{Schematic diagram of an \textit{IUCNN} workflow for fully connected neural network models. The relevant \textit{IUCNN} functions are shown in the grey boxes. Symbol picture of species distribution from www.gbif.org}
\end{figure}
library(tidyverse)
library(IUCNN)

# Load occurrences for training species and getting features
train_occ <- read.csv("S2_training_occurrences_example.csv")
train_feat <- iucnn_prepare_features(train_occ, type = "geographic")

# Load and prepare training labels
rl_labels <- read.csv("S2_training_labels_example.csv")
train_lab <- iucnn_prepare_labels(rl_labels,
    train_feat,
    level = "detail")

# Load occurrences for prediction species and get features
pred_occ <- read.csv("S2_prediction_occurrences_example.csv")
pred_feat <- iucnn_prepare_features(pred_occ, type = "geographic")

# Model testing
# For illustration models that differ in dropout rate and number of layers
mod_test <- iucnn_modeltest(train_feat,
    train_lab,
    logfile = "model_testing_results.txt",
    model_outpath = "iucnn_modeltest",
    mode = "nn-class",
    dropout_rate = c(0.0, 0.1, 0.3),
    n_layers = c("30", "40_20", "50_30_10"),
    cv_fold = 5,
    init_logfile = TRUE)

# Select best model
m_best <- iucnn_best_model(mod_test,
    criterion = "val_acc",
    require_dropout = TRUE)

# Inspect model structure and performance
summary(m_best)
plot(m_best)

# Train the best model on all training data for prediction
m_prod <- iucnn_train_model(train_feat,
    train_lab,
    production_model = m_best,
    overwrite = TRUE)

# Predict RL categories for target species
pred <- iucnn_predict_status(pred_feat,
    m_prod)
plot(pred)

FIGURE 2  Example code on how to train and test a fully connected neural network using IUCNN and use it to predict the RL status of 13,207 orchid species, including model testing and selection ("Option 2" in Figure 1)
on any trait that can be summarized at the species level and might be informative of species RL category, including, for example, phenotypic traits or information about human use of the species. Yet, because for many species only geographic occurrence data are readily available, we implemented the \texttt{iucnn\_prepare\_features} function to prepare default features from user-provided geo-referenced occurrence records. The function automatically downloads publicly available environmental data and matches these to the user-provided occurrences. For terrestrial taxa, the resulting default features include characteristics of the species range, the human impact within the species range (Venter et al., 2016, https://wcshumanfotprint.org), the climatic conditions across a species range (Fick & Hijmans, 2017, https://www.worldclim.org) and species occurrences in different biomes (Olson et al., 2001, http://assets.worldwildlife.org). We provide a detailed list of all features in Table S1 in Appendix S1. If no custom dataset of geo-referenced species occurrences is available, these can be obtained from public databases, for instance, the Global Biodiversity Information Facility (GBIF). If data from an online database are used, taxonomic scrubbing (Cayuela et al., 2012; Freiberg et al., 2020), quantification of sampling bias (Zizka et al., 2021) and geographic cleaning (Arlé et al., 2021; Zizka et al., 2019) are advisable (Maldonado et al., 2015; Zizka et al., 2020).

Differences in sampling effort among regions (geographic sampling bias) can undermine automated conservation assessment. For instance, if an automated conservation assessment is exclusively based on public available records and geographic sampling bias is strong (Hughes et al., 2021), species of sparsely sampled regions can receive overly alarmist assessments. This effect has been demonstrated for automated assessment methods that are based on species range size alone (such as ConR, rCAT and SPCG included in our empirical example), which will classify any species with <15 occurrence records as potentially threatened (Lughadha et al., 2019; Rivers et al., 2011). We expect the neural networks implemented in \texttt{IUCNN} to be more robust to geographic sampling bias because features such as longitude and latitude as well as latitudinal and longitudinal range inform the model on how to weigh the number of available occurrence records in different regions. Additionally, users can design features to address explicitly geographic sampling bias for instance with the \texttt{iucnn\_bias\_feature} function in \texttt{IUCNN}. This function uses \texttt{sampbias} (Zizka, Antonelli, et al., 2021) to calculate the expected sampling bias dependent on human accessibility across species’ ranges as feature for the \texttt{IUCNN} models (currently only feasible for the local to regional scale). Accounting for geographic sampling bias in biodiversity science is an active field of research. The high flexibility of neural networks in \texttt{IUCNN} will make it easy for users to apply different approaches to factor geographic sampling biases into automated conservation assessments, for instance, by scrutinizing data structure (Boyd et al., 2021), spatial thinning (Aiello-Lammens et al., 2015) and modelling (Bruehlheide et al., 2020; Varela et al., 2014).

Users may provide a dataset of custom features, if available. Features may be continuous or discrete and can represent any trait considered relevant for approximating the conservation status in the region and taxon of interest, such as morphological and phenotypic data or population dynamics. Features should be rescaled so that their numeric values range in similar orders of magnitude to facilitate model convergence, and missing data should be coded with distinct values or imputed. When using custom features, it is important that the same features are provided in the training and prediction datasets, and that they are rescaled and ordered consistently.

\texttt{IUCNN} contains the \texttt{iucnn\_prepare\_labels} function to prepare standardized labels (i.e. the five IUCN categories, or a binary classification: Possibly threatened/Not threatened) in the format needed for model training based on a table of species names and RL status assessments, and a feature set prepared with \texttt{iucnn\_prepare\_features} (to standardize species order, Figure 2).

### 2.2 Model training and validation

We implemented four models in \texttt{IUCNN}, which we term: \texttt{cnn-class}, \texttt{nn-class}, \texttt{bnn-class} and \texttt{nn-reg}.

The \texttt{cnn-class} model is based on a CNN which is a deep learning model typically used to process patterns for classification, for instance, for image identification. In \texttt{IUCNN}, this model has the advantage of extracting features from the raw (rasterized) data as part of the training, unlike the fully connected NN models that rely on predefined features. Thus, the \texttt{cnn-class} model is less dependent on the user choices as it learns from the patterns observed in the data. This comes, however, with the disadvantage of additional parameters (in the convolutional part of the model), which in turn means that a larger training set may be required to optimize the model. The other drawback compared to the fully connected NN models implemented in \texttt{IUCNN} is that additional traits cannot as easily be incorporated into the predictions. Yet, in our empirical analyses on the orchid dataset, the \texttt{cnn-class} model yielded a similar accuracy as the other models (Figure 3). While further empirical applications will be needed to explicitly test the generality of this result, we expect the \texttt{cnn-class} model to perform particularly well in regional rather than global assessments, given that it can learn how to map spatial patterns into a threat prediction.

In the fully connected NN models (\texttt{nn-class}, \texttt{nn-reg} and \texttt{bnn-class}), the input layer consists of a set of quantitative and categorical features computed for each species. The input features are mapped onto the output layer through one or more hidden layers.

The \texttt{nn-class} model uses a SoftMax activation function in the output layer to obtain a vector of probabilities, with one value for each class, for example, the five RL categories. The prediction under this model is determined by the class that received the largest probability value. The probabilities are then used to compute the cross-entropy loss, which is averaged across all training instances and minimized during the optimization. The \texttt{nn-reg} model predicts the conservation status of species as a regression task. The output layer of the \texttt{nn-reg} model consists of a single value, which may be taken as is, or transformed through a sigmoid or tanh activation function (depending on user settings). Next, the output value is compared with the original or rescaled RL status during model optimization as mean
FIGURE 3 The performance of different IUCNN models trained on 886 species with existing RL assessments at www.iucnredlist.org. (a) The overall test accuracy of different automated methods to estimate the extinction risk of species on the level of RL categories or the broader Possibly threatened/Not threatened level. ConR, rCat, and SPGC are other automation methods for comparison. (b) Confusion matrices of the three IUCNN models, showing the number of correctly predicted instances for each category, as produced by cross-validation at the detailed RL category. (c) As b for the broad Possibly threatened/Not threatened level. Numbers font color in black or white for readability.
squared error (MSE). IUCNN then transforms the predicted value into a categorical prediction by rounding the output to the closest class. Hence, the nn-reg model explicitly incorporates the ordinal nature of the RL categories. The potential advantage of this approach is that the error is weighted by the degree of disagreement between the truth and the prediction. For instance, the distance (error) between NT and LC is considered much smaller than between CR and LC in nn-reg. Thus, while the accuracies are expected to be similar among the three full connected NN models, the magnitude of the error in misclassified instances is expected to be smaller in the nn-reg model. The nn-class and nn-reg models are built using the Tensorflow module (v2 or greater; Abadi et al., 2015) and use the Adam gradient descent for optimization (Kingma & Ba, 2017).

The bnn-class model is a Bayesian implementation of a NN classifier using the npBNN module (https://github.com/dsilvestro/npBNN). In the bnn-class model, the output class probabilities are used as the parameters of a categorical probability mass function to compute the likelihood of the training data. It is similar in structure to the nn-class model with the advantage of producing a posterior sample of predictions for each instance instead of a point estimate. From these samples, we compute posterior probabilities associated with each class, thus providing a direct estimation of the uncertainty around the prediction. As demonstrated in the empirical example, the posterior probabilities can be used to determine a threshold above which instances are expected to be classified with an accuracy matching or exceeding a user-defined target. The advantage of the BNN implementation comes at the cost of a time-consuming optimization. The bnn-class model is trained through a Markov Chain Monte Carlo (MCMC) algorithm, sampling the weights from their posterior distribution. By default, normal density functions with mean equal to 0 and standard deviation set to 1 are used as prior distributions applied to the weight parameters. To also quantify the uncertainties of the predictions made by the nn-class and nn-reg models, we implemented the Monte Carlo dropout method (Gal & Ghahramani, 2016). While the resulting dropout probabilities are not posterior probabilities, they can be similarly applied and interpreted as a measure of prediction uncertainty and allow users to determine for which species a prediction can be made with a defined level of confidence.

The training of the fully connected NN models is implemented in the iucnn_train_class function (Figures 1 and 2), while for the cnn-class model this is done using the iucnn_train_cnn function due to the differing feature structure. Models can be trained without prior knowledge about neural networks using the default. For advanced users, the functions provide arguments to adjust model parameters (e.g. the number of hidden layers, the number of nodes per layer and the activation functions) and to customize the training process (e.g. the number of epochs, stopping criteria for nn-class and nn-reg and the number of MCMC iterations and the sampling frequency for bnn-class). The performance of trained IUCNN models can be evaluated using the summary function, which calculates summary statistics for the model on unseen data including an overall prediction accuracy and a confusion matrix. Additionally, the plot function can be used to plot the loss of the trained IUCNN model throughout the training epochs (nn-class and nn-reg) or the posterior samples (bnn-class), which helps to evaluate whether or not the model has converged. To evaluate a range of different model settings, we implemented the iucnn_modeltest function, which allows the user to choose among different NN architectures and hyperparameter configurations (e.g. number of layers, nodes per layer and activation functions) using a cross-validation approach within nn-class and nn-reg models. Finally, users can evaluate how much the chosen model relies on the different types of features using the iucnn_feature_importance function (not available for CNN models). The implemented process of permutation feature importance evaluates the loss of prediction accuracy when the signal in individual features is muted by randomly shuffling the feature values among instances. The resulting feature importance values for each feature or block of features can be plotted with the plot function. This can help users to decide which features are most important and should be included in a model to obtain the best prediction accuracy (Figure 1; Table 1). We provide a tutorial on how to train and test models as a vignette that comes with IUCNN (Appendix S2).

2.3 Prediction and visualization

Provided with the trained model and the features for the target species, the iucnn_predict_status function predicts RL categories for these species. Labels may be returned as a vector of labels or optionally, in the case of an nn-reg analysis, as the raw regressed values. In case a target accuracy was set, all species for which the RL category cannot be predicted at the desired accuracy are labelled as NA. Using the plot function, users can plot a histogram of the numbers of predicted species per category.

2.4 Implementation

IUCNN is implemented as an R-package (R Core Team, 2021) with integrated Python code. All user-level functions are accessible via R, since this language is widely used in ecological and conservation research, and no Python knowledge is required to use IUCNN. We developed IUCNN using the “usethis” package (Wickham & Bryan, 2020) following Wickham and Bryan (2021).

IUCNN depends on dplyr (Wickham et al., 2019), magrittr (Bache & Wickham, 2014), readr (Wickham & Hester, 2020), tidyr (Wickham, 2020) and tidyselect (Henry & Wickham, 2020) for data wrangling; raster (Hijmans, 2018), rCAT (Moat, 2017), sf (Pebesma, 2018), stats and terra (Hijmans, 2021) for feature preparation; graphics and grDevices for visualization; and reticulate (Ushey et al., 2020) for integrating R and Python. Furthermore, IUCNN suggests checkmate (Lang, 2017), covr (Hester, 2020), spelling (Ooms & Hester, 2020) and testthat (Wickham, 2011) to secure code functionality; and knitr (Xie, 2020) and pkgdown (Wickham & Hesselberth, 2020) for documentation.
All software needed to run IUCNN can be installed with a few lines of code. The current version of IUCNN can be installed from GitHub (https://github.com/IUCNN/IUCNN), from within R using the devtools (Wickham et al., 2020) package. Since the neural networks are trained and used in Python, Python also needs to be installed, including the tensorflow and npBNN modules. This can be done from within R using the reticulate package (Ushey et al., 2020). The readme file on IUCNN’s GitHub page and the vignette provided with the package provide the instructions for installation.

### 2.5 | Empirical example

We demonstrate the use of IUCNN using an empirical dataset on the global distribution and extinction risk of orchid species (Orchidaceae). The dataset contains 14,093 orchid species from across the globe, 886 of them (c. 6%) with existing RL assessments (the training data). The specific dataset is described in detail in Zizka, Silvestro, et al. (2021), and the occurrence records are originally obtained from GBIF (Global Biodiversity Information Facility; www.gbif.org, 2019). In Zizka, Silvestro, et al., 2021, the dataset has already been analysed with the ConR (Dauby et al., 2017), rCAT (Moat, 2017) and SPGC (Schmidt et al., 2017; Töpel et al., 2017) automation methods which we use for comparison. These three methods are comparable to IUCNN in that they can be solely based on species geographic occurrence records and are implemented in accessible R-packages, but differ in that they do not predict extinction risk based on traits, but calculate RL indices (Extent of Occurrence and Area of Occupancy) relevant for red listing under Criterion B, and may be interpreted as preliminary assessment.
To illustrate the use and flexibility of IUCNN, we train \( \text{cn-class} \), \( \text{nn-class} \), \( \text{nn-reg} \) and \( \text{bn-class} \) models based on different sets of features, with different model structures and two different levels of prediction detail (five RL categories, and Possibly threatened vs. Not threatened species, respectively). We tested \( \text{bn-class} \), \( \text{nn-class} \) and \( \text{nn-reg} \) models based on geographic features, geographic plus human footprint features and “all features.” “All features” included all default features available in IUCNN (geographic, biome, climate and human footprint) plus a custom feature related to orchid growth form, which we included to illustrate the use of phenotypic traits in IUCNN analyses. Growth form – epiphytic vs. not-epiphytic – is a well-known axis of ecological differentiation in orchids also affecting traits relevant for conservation and threat estimation. For instance, epiphytic species depend on large woody host plants and predominantly occur in tropical regions. We use a binary feature epiphytic (y/n), and classified all species as epiphytic listed in the global Epilist v1.0 database (Zotz et al., 2021).

We compared different \( \text{nn-class} \) models with 1, 2 and 3 hidden layers and dropout rates set to 0, 0.1 and 0.3 (detailed settings are available in Appendix S2). For \( \text{nn-reg} \) models, we tested the same specifications and additionally tested different output layer activation functions. We tested all combinations of these settings. Within each model type and output classification, we chose the best model based on the cross-validation accuracy. Due to the longer convergence time, we ran the \( \text{bn-class} \) and \( \text{cn-class} \) models assuming a single configuration. We used a \( \text{bn-class} \) model with one hidden layer with 20 nodes and a \( \text{nn-class} \) model with two convolutional layers (with 1 filter and kernel size of \( 3 \times 3 \)) followed by average pooling (with pool size of \( 3 \times 3 \)) after preliminary testing. In the case of \( \text{cn-class} \), the flattened output was then fed into a fully connected network of two hidden layers (with 20 and 10 nodes with ReLU activation functions) with a dropout layer after each hidden layer and a dropout rate of 10%. As for the other classification models, we then used an output layer with SoftMax activation function.

The overall accuracy of models was similar across the tested subsets of input features. In general, models based solely on geographic features performed best (in 4 of 6 cases, more often than any other feature combination, Figure S1 in Appendix S1), and we therefore used these models for further analyses.

The accuracy of any IUCNN model was higher than the accuracy of the available automation methods based on Criterion B (Figure 3a). For all IUCNN models, the overall accuracy was higher at the Possibly threatened vs. Not threatened level (\( \text{cn-class} \): 0.77, \( \text{nn-class} \): 0.81, \( \text{nn-reg} \): 0.80 and \( \text{bn-class} \): 0.80; Figure 3a) than at the detailed level with all five RL categories (\( \text{cn-class} \): 0.52, \( \text{nn-class} \): 0.60, \( \text{nn-reg} \): 0.54 and \( \text{bn-class} \): 0.60; Figure 3a). While the overall accuracy was slightly higher for the \( \text{nn-class} \) and \( \text{bn-class} \) models, the \( \text{nn-reg} \) and \( \text{cn-class} \) models performed better in including intermediate categories, in particular NT and VU (Figure 3b). Given that most species fall into the LC category, this advantage of the \( \text{nn-reg} \) model was not reflected in the overall prediction accuracy. When using target accuracy thresholds, the proportion of species evaluated decreased with increasing target accuracy for all model types and detail levels (Figure 4). With increasing target accuracy, species of intermediate categories are subsequently removed and species that can be classified at higher accuracies mostly belong to the extreme LC and CR categories (Figure S2 in Appendix S1). In the case of the best \( \text{nn-class} \) model with five RL classes, a target accuracy of 85% retains 3,310 species (of 13,207), 3,004 of them as LC, 104 as EN and 202 as CR.

### 3 | DISCUSSION

We presented IUCNN, a user-friendly R-package, to use neural networks for the prediction of species RL extinction risk assessments. IUCNN is flexible with regard to input data and may be used solely based on publicly available information. Our empirical example using more than 13,000 orchid species demonstrates that the models implemented in IUCNN can be run with few lines of code and can outcompete other automated methods in overall accuracy.

Predictive models, such as the neural networks implemented in IUCNN, provide approximations of species RL assessments, which can be used in biogeographic and (macro)ecological research. Specifically, these approximations are suitable to broaden the taxonomic and geographic scope of synthetic research into understanding the distribution and mechanisms of threat and extinction risk. When extinction risk is used, for instance, to understand the geographic distribution of threatened species or to quantify the impact of species traits on extinction risk, predictive approaches can overcome taxonomic and geographic biases, and enable such analyses for organisms that are poorly represented in the current RL. Furthermore, predictive approaches can support ecological, biogeographic and taxonomic case studies with information on species extinction risk. For example, added information on the estimated number and identity of threatened species in a study taxon or region may increase the conservation relevance of research, help to set research priorities or even guide ethical decisions on sampling strategies (for instance, by switching to less invasive sampling techniques for Possibly threatened species).

We designed IUCNN for user-friendly access and included multiple features to address common issues arising with the use of RL data for prediction of species extinction risk. Four specific strengths of IUCNN are as follows:

#### 3.1 | Easy model evaluation and model testing

Careful model evaluation and testing is crucial for all statistical models. IUCNN provides options for model evaluation, testing and customization. When predicting extinction risk and training on RL data, high accuracies are crucial and uncertainty must be clearly visible (since errors may result in a misprioritization of conservation resources). The implementation of Bayesian neural networks allows for a probabilistic quantification of uncertainty, and the target accuracy threshold allows limiting prediction to high levels of accuracy.
The use of cross-validation for model testing and the option to subsequently train a model on all available data ensure to use as much of the existing training data as possible to address a limited amount of training data.

### 3.2 | High accessibility

*IUCNN* is implemented in R and an individual analysis may be run with few lines of code in a straightforward workflow (Figures 1 and 2). *IUCNN* contains utility functions to prepare standardized features in the required input format suggesting a set of default features that have proven relevant in empirical analyses using fully connected NNs (iucnn_prepare_features) and CNNs (iucnn_cnn_features) as well as easy options to customize them via separate functions (iucnn_bias_features, iucnn_biome_features, iucnn_climate_features, iucnn_footprint_features and iucnn_geography_features). Model testing and evaluation are concentrated in just three functions (iucnn_modeltest, iucnn_best_model and iucnn_feature_importance) and the results can easily be summarized (summary) and visualized (plot). All functions are documented at the standards of the Comprehensive R Archive Network and a detailed vignette is part of the package. Furthermore, the majority of *IUCNN* analyses (including several thousand species and hundreds of features) can be run on a standard laptop.

### 3.3 | High flexibility

*IUCNN* may predict species extinction risk based on any traits deemed relevant by the users, as long as they can be summarized at the species level (in the case of the NN models). While we provide custom features that can be derived from data publicly available for many taxa, users can easily provide additional data describing, for instance, the physiology, taxonomy, morphology and ecology of species. With CNN-based models, features, although currently limited to spatial patterns, do not even need to be specified by the user but are instead extracted as part of the learning process from the raw occurrence data directly. This flexibility can avoid circularities when, for example, using *IUCNN* predictions to relate extinction risk to specific traits. Furthermore, *IUCNN* may use any user-defined output labels. These may be the default five IUCN categories (LC, NT, VU, EN and CR) but may also be any user-defined classes, such as the binary classification, demonstrated in the empirical example.

---

**FIGURE 4** The effect of different user-defined targets for overall model accuracy on the number of species assessed. Numbers are based on the empirical dataset of 13,207 orchid species with no official assessment available. The number of assessed species decreases with increasing accuracy, but in this example the nn-class model can classify more than 3,000 species at 85% accuracy. Left panel: five output classes (IUCN categories: CR, EN, VU, NT, LC). Right panel: two output classes (Possibly threatened, Not threatened).
or for instance, classification schemes of regional Red Lists, which often are different.

3.4 | Improved accuracy

Neural networks can outcompete comparable methods in identifying Possibly threatened species and classifying species into RL categories, as demonstrated by the empirical analysis (Figure 3a). Neural networks and deep learning are an active field of research, with multiple lines of development on the horizon. Together with the growing amount of RL data that can be used for training and the increase in remote sensing technologies for feature preparation, we anticipate an increase in application cases and accuracy of IUCNN models.

3.5 | Future prospects

For any IUCNN analysis, a careful choice of input data is crucial. In general, the performance of supervised learning models increases with the amount of training data (Goodfellow et al., 2016). In particular, we recommend choosing training taxa and regions to comprise at least several hundred taxa (885 in our example) and, when possible, with a good representation of all RL classes. The standardization of RL criteria across taxa (IUCN Standards & Petitions Subcommittee, 2017) favours the choice of broad reference taxa and makes IUCNN equally suited for plant and animal species and regions worldwide. If marine or freshwater species are the primary focus, feature design should account for particularities of these systems, for instance, the linearity of river systems and depth preferences of marine species (Bland, 2017). If occurrence data from museums or herbaria are the basis for feature design, record age can be of concern. First, because the geographic information of older specimen is usually less precise, which we recommend addressing during data cleaning (Zizka et al., 2019). Second, because old records can represent occurrences in areas where species do no longer occur due to climate or land-use change. One way to inform the model on this temporal change is to combine features calculated based on different time slices of input data.

Declines in species ranges and populations are important for red listing (particularly criteria A and D). If sufficient data are available on temporal changes in species range and population size, those can be included into IUCNN analyses. Users may design features that directly represent temporal trends (for instance, population counts at different time points or changes in population size). If such data are not available, temporal changes can be included via features resulting from temporal stratification of occurrence records (for instance, by calculating the Extent of Occurrence between 1980 and 1990 and the extent of occurrence between 2005 and 2015). Additionally environmental data can be temporally stratified (for instance, the fraction of species range covered with forest between 1980 and 1990 and 2005 and 2010). Such temporally stratification of environmental features is implemented in IUCNN for the human footprint index, via the iucnn_footprint_features function. It should be noted that if records are split by time intervals, it may be relevant to explicitly consider temporal sampling bias (differences in sampling effort at different times; e.g. Boyd et al., 2021; Zizka, Rydén, et al., 2021).

While we are confident that the specific adaptations of IUCNN make it a useful tool for research, we emphasize that trait-based predictions of species extinction risk are only approximations and need to be interpreted with caution (Walker et al., 2020). As for any statistical model, the quality of IUCNN predictions will depend on the quality of the input data and model fit. In the specific RL framework, the structure of the RL likely limits the maximum accuracy that can be achieved in predicting species extinction risk based on traits. The RL is a collaborative effort of many different specialist groups and professional assessors generating the assessments for individual taxonomic groups. Although assessments are based on a standardized framework, different assessors may interpret criteria and categories differently, especially the less clearly defined NT and VU categories. This effect is exacerbated by the differences in data availability among taxa and the resulting varying reliability of individual assessments. Furthermore, species may be listed on the RL based on different criteria, for instance, either geographic range (Criterion B) or population trends (Criterion A), or both, and assessors include expert knowledge and evaluation into the assessment. Hence, any set of input features may only incompletely capture the red-listing process and therefore misclassify species.

A specific caveat of predicting RL categories is their imbalanced distribution. In our empirical example, the models accurately identified species in the extreme categories (LC and CR), but performed less well with intermediate categories – the majority of misclassifications were related to NT and VU categories (Figure 3b). This is likely related to these categories being relatively rare on the RL and thus underrepresented in the training data, which may bias predictions to reproduce category frequencies observed in the input data. Indeed, the models used for the empirical data were biased towards class label frequencies in the training data, especially if the frequencies were very different between training and test sets (Figure S3 in Appendix S1). The implications of class imbalance for model application will depend on the expected similarity in class frequencies among species already evaluated on the RL (the training data) and DD and NE species (the target species), as well as the modelling algorithm (Figure S3). Conceptually, the expected difference in category frequencies remains unclear because different factors affect the class frequencies of RL categories in a given set of species. On the one hand, by design, most species in any dataset will be LC, often also among the DD and NE species (Butchart & Bird, 2010). On the other hand, compared to species on the RL, datasets of DD and NE species are likely to comprise a higher proportion of rare species, which may be more threatened (Parsons, 2016). Yet, DD and NE species will more often occur in regions difficult to access for IUCN assessors (Bland et al., 2017) and might also be subject to reduced human pressure in general, and therefore have a reduced extinction risk. We acknowledge these conceptual and practical limitations for RL...
status predictions in general and have therefore included a battery of options to address these issues in IUCNN. These include the \texttt{nnreg} models to account for the ordinal nature of the RL categories, the flexible detail levels in the prediction and the target accuracy threshold to reduce uncertainty.

In conclusion, IUCNN is a user-friendly implementation of deep learning methods to approximate RL extinction risk assessments for species that are so far Data Deficient or Not Evaluated on the RL. IUCNN provides new tools to estimate species extinction risk and makes these innovative methods available to a larger community.

\section*{Acknowledgements}

We thank the IUCN and all Red List assessors as well as the data collectors and data providers to GBIF for their effort. We are thankful to Steven Bachman and Barnaby Walker for helpful discussion during the development of IUCNN. AZ was funded by iDiv via the German Research Foundation (DFG FZT 118), specifically through sDiv, the Synthesis Centre of iDiv. TA and DS received funding from the Swedish Research Council (VR: 2019-04739). DS received funding from the Swiss National Science Foundation (PCEFP3_187012; FN-1749). Open Access funding enabled and organized by Projekt DEAL.
Rapacciuolo, G. (2019). Strengthening the contribution of macro-ecological models to conservation practice. *Global Ecology and Biogeography, 28*(1), 54–60. https://doi.org/10.1111/geb.12848

Richards, C., Cooke, R. S. C., & Bates, A. E. (2021). Biological traits of seabirds predict extinction risk and vulnerability to anthropogenic threats. *Global Ecology and Biogeography, 30*(5), 973–986. https://doi.org/10.1111/geb.13279

Rivers, M. C., Taylor, L., Brummitt, N. A., Meagher, T. R., Roberts, D. L., & Lughadha, E. N. (2011). How many herbarium specimens are needed to detect threatened species? *Biological Conservation, 144*(10), 2541–2547. https://doi.org/10.1016/j.biocon.2011.07.014

Rolland, J., & Salamin, N. (2016). Niche width impacts vertebrate diversification. *Global Ecology and Biogeography, 25*(10), 1252–1263. https://doi.org/10.1111/geb.12482

Rondinini, C., Marco, M. D., Visconti, P., Butchart, S. H. M., & Boitani, L. (2014). Update or outdate: Long-term viability of the IUCN Red List. *Conservation Letters, 7*(2), 126–130. https://doi.org/10.1111/conl.12040

Santini, L., Antão, L. H., Jung, M., Benitez-López, A., Rapacciuolo, G., Di Marco, M., Jones, F. A. M., Haghkerdar, J. M., & González-Suárez, M. (2021). The interface between macroecology and conservation: Existing links and unlinked opportunities. *Frontiers of Biogeography, 13*(4), e53025. https://doi.org/10.21425/FSFBG53025

Schmidt, M., Zizka, A., Traoré, S., Ataholo, M., Chatelain, C., Daget, P., Dressier, S., Hahn, K., Kirchmair, I., Krohmer, J., Mbayongne, E., Müller, J. V., Nacoulma, B., Ouédraogo, A., Ouédraogo, O., Sambaré, O., Schuman, K., Wieringa, J. J., Zizka, G., & Thiombiano, A. (2017). Diversity, distribution and preliminary conservation status of the flora of Burkina Faso. *Phytotaxa Monographs, 304*(1), 1–215. https://doi.org/10.11646/phytotaxa.304.1.1

Smiley, T. M., Title, P. O., Zelditch, M. L., & Terry, R. C. (2020). Multidimensional biodiversity hotspots and the future of taxonomic, ecological and phylogenetic diversity: A case study of North American rodents. *Global Ecology and Biogeography, 29*(3), 516–533. https://doi.org/10.1111/geb.13050

Tingley, R., Mahoney, P. J., Durso, A. M., Tallian, A. G., Moran-Ordonez, A., & Beard, K. H. (2016). Threatened and invasive reptiles are not two sides of the same coin. *Global Ecology and Biogeography, 25*(9), 1050–1060. https://doi.org/10.1111/geb.12462

Töpel, M., Zizka, A., Caliò, M. F., Scharn, R., Silvestro, D., & Antonelli, A. (2017). SpeciesGeoCoder: Fast categorization of species occurrences for analyses of biodiversity, biogeography, ecology, and evolution. *Systematic Biology, 66*(2), 145–151. https://doi.org/10.1093/sysbio/syw064

Ushey, K., Allaire, J. J., & Tang, Y. (2020). *reticulate: Interface to ‘Python*. Retrieved from https://CRAN.R-project.org/package=reticulate

Varela, S., Anderson, R. P., García-Valdés, R., & Fernández-González, F. (2014). Environmental filters reduce the effects of sampling bias and improve predictions of ecological niche models. *Ecography, 37*(11), 1084–1091. https://doi.org/10.1111/1600-0587.12044.x

Venter, O., Sanderson, E. W., Magrach, A., Allan, J. R., Beher, J., Jones, K. R., Possingham, H. P., Laurance, W. F., Wood, P., Fekete, B. M., Levy, M. A., & Watson, J. E. M. (2016). Global terrestrial human footprint maps for 1993 and 2009. *Scientific Data, 3*(1), 160067. https://doi.org/10.1038/sdata.2016.67

Walker, B. E., Leão, T. C. C., Bachman, S. P., Bolam, F. C., & Nic Lughadha, E. (2020). Caution needed when predicting species threat status for conservation prioritization on a global scale. *Frontiers in Plant Science, 11*(April), 1–4. https://doi.org/10.3389/fpls.2020.00520

Walsh, J. C., Venter, O., Watson, J. E. M., Fuller, R. A., Blackburn, T. M., & Possingham, H. P. (2012). Exotic species richness and native species endemism increase the impact of exotic species on islands. *Global Ecology and Biogeography, 21*(8), 841–850. https://doi.org/10.1111/j.1466-8238.2011.00724.x

Wickham, H. (2011). testthat: Get Started with Testing. *The R Journal, 3*, 5–10. https://doi.org/10.32614/RJ-2011-002

Wickham, H. (2020). *tidy: Tidy Messy Data*. Retrieved from https://CRAN.R-project.org/package=tidy

Wickham, H., & Bryan, J. (2020). *usethis: Automate Package and Project Setup*. Retrieved from https://CRAN.R-project.org/package=usethis

Wickham, H., & Bryan, J. (2021). *R Packages (2nd ed.)*. O’Reilly. Retrieved from https://r-pkgs.org/index.html

Wickham, H., François, R., Henry, L., & Müller, K. (2021). *dplyr: A Grammar of Data Manipulation*. Retrieved from https://CRAN.R-project.org/package=dplyr

Wickham, H., & Hesselberth, J. (2020). *pkgdown: Make Static HTML Documentation for a Package*. Retrieved from https://CRAN.R-project.org/package=pkgdown

Wickham, H., & Hester, J. (2020). *readr: Read rectangular text data*. Retrieved from https://CRAN.R-project.org/package=readr

Wickham, H., Hester, J., & Chang, W. (2020). *devtools: Tools to Make Developing R Packages Easier*. Retrieved from https://CRAN.R-project.org/package=devtools

Xie, Y. (2020). *knitr: A general-purpose package for dynamic report generation*. Retrieved from https://yihui.org/knitr/

Zizka, A., Antonelli, A., & Silvestro, D. (2021). Sampbias, a method for quantifying geographic sampling biases in species distribution data. *Ecography, 44*(1), 25–32. https://doi.org/10.1111/ecog.05102

Zizka, A., Antunes Carvalho, F., Calvente, A., Cabral, A., Coelho, J. F. R., Colli-Silva, M., Fantinati, M. R., Fernandes, M. F., Ferreira-Araújo, T., Gondim Lambert Moreira, F., Santos, N. M. C., Santos, T. A. B., dos Santos-Costa, R. C., Serrano, F. C., Álves da Silva, A. P., de Souza Soares, A., Cavalcante de Souza, P. G., Calisto Tomaz, E., ... Antonelli, A. (2020). No one-size-fits-all solution to clean GBIF. *PeerJ, 8*, e9916. https://doi.org/10.7717/peerj.9916

Zizka, A., Barratt, C. D., Duarte Ritter, C., Joerger-Hickfang, T., & Zizka, V. M. A. (2021). Existing approaches and future directions to link macroecology, macroevolution and conservation prioritization. *Ecography. doi.org/10.1111/ecog.05557

Zizka, A., Rydén, O., Edler, D., Klein, J., Perrigo, A., Silvestro, D., Jagers, S. C., Lindberg, S. I., & Antonelli, A. (2021). Bio-Dem, a tool to explore the relationship between biodiversity data availability and socio-political conditions in time and space. *Journal of Biogeography, 48*(11), 2715–2726.

Zizka, A., Silvestro, D., Andermann, T., Azevedo, J., Duarte Ritter, C., Edler, D., Farooq, H., Herdean, A., Ariza, M., Scharn, R., Santesson, S., Wengström, N., Zizka, V. M. A., & Antonelli, A. (2019). *CoordinateCleaner: Standardized cleaning of occurrence records from biologically relevant collection databases*. Methods in Ecology and Evolution, 10(5), 744–751. https://doi.org/10.1111/2041-210X.13152

Zizka, A., Silvestro, D., Witt, P., & Knight, T. M. (2021). Automated conservation assessment of the orchid family with deep learning. *Conservation Biology, 35*(3), 897–908. https://doi.org/10.1111/cobi.13616

Zotz, G., Weigelt, P., Kessler, M., Kreft, H., & Taylor, A. (2021). *EpiList 1.0: A global checklist of vascular epiphytes*. *Ecology, 102*(6), e03326. https://doi.org/10.1002/ecy.3326
AZ is a botanist and biogeographer dedicated to developing computational methods for macroecology, macroevolution and conservation of biodiversity. TA is a computational biologist and bioinformatician, developing machine learning solutions for biodiversity research. DS is a computational biologist developing models and software in evolutionary and conservation biology.

Author contribution: AZ and DS conceived of this study. All authors developed the analyses workflows and designed the NN models. AZ implemented the data processing and TA the machine learning functions of the IUCNN tool, with contributions from DS. All authors analyzed the empirical data. AZ wrote the manuscript with contributions from DS and TA.