Ecological Informatics 94 (2026) 103657

Contents lists available at ScienceDirect

Ecological Informatics

journal homepage: www.elsevier.com/locate/ecolinf

# PNW-Cnet: An evolving convolutional neural network to support broad-scale passive acoustic monitoring

Damon B. Lesmeister [a,b,*], Zachary J. Ruff [a,b], Adam Duarte [c], Anna B. Kohlberg [a,b], Taal Levi [b], Christopher M. Sullivan [d]

[a] USDA Forest Service, Pacific Northwest Research Station, Corvallis, OR 97331, USA
[b] Department of Fisheries, Wildlife, and Conservation Sciences, Oregon State University, Corvallis, OR 97331, USA
[c] USDA Forest Service, Pacific Northwest Research Station, Olympia, WA 98512, USA
[d] College of Earth, Ocean, and Atmospheric Sciences, Oregon State University, Corvallis, OR, USA

A R T I C L E I N F O

*Keywords:*
Passive acoustic monitoring
Deep learning model
Soundscape ecology
Species detection
Forest ecosystems
Wildlife monitoring

A B S T R A C T

Passive acoustic monitoring is increasingly used for broad-scale, noninvasive tracking of biodiversity across space and time, but processing the resulting audio data at scale remains a major challenge. To address this, we have iteratively developed PNW-Cnet, a convolutional neural network trained to detect wildlife vocalizations in unprocessed field recordings collected throughout a large, forested region of the western United States. Initially built to identify six owl species, PNW-Cnet has evolved through successive retrainings and dataset expansion. We trained PNW-Cnet v5 on 824,120 labeled spectrograms to detect 135 sonotypes (hereafter, classes), including birds (72 species), mammals (11 species), and one amphibian species, as well as diverse anthropogenic and environmental sounds. We also implemented a targeted data augmentation strategy to address class imbalance, enabling the inclusion of rare or underrepresented species and noise types in model training. Here we describe the architecture and training of PNW-Cnet v5, its integration with a user-friendly Python package (pycnet-audio) and a Shiny-based graphical interface for streamlined data processing, and its performance across a range of classes. At a 0.95 confidence score detection threshold, the model achieved $\geq$95% precision for 112 target classes in real-world applications. Precision remained high across lower confidence score thresholds, while recall varied with class representation in the training set and with vocalization characteristics. PNW-Cnet exemplifies how deep learning models can be recursively expanded and operationalized to support real-world endangered species population monitoring, acoustic community characterization, disturbance mapping, and ecological monitoring. This approach enables the rapid detection of target classes within audio recordings, unlocking new potential for passive acoustic monitoring in biodiversity science and conservation management.

## 1. Introduction

Monitoring biodiversity at broad spatial and temporal scales is essential for understanding ecosystem change, assessing conservation outcomes, and informing management decisions in the face of intensifying environmental pressures (Anderson, 2018; Pereira et al., 2013). Traditional wildlife monitoring approaches (e.g., visual encounter surveys, point counts, and live trapping) often require extensive personnel time, specialized expertise, and physical access to survey locations, which can constrain their spatiotemporal application (Lindenmayer and Likens, 2010; Yoccoz et al., 2001). These limitations are especially pronounced for cryptic, nocturnal, or rare species, and in ecosystems that are remote, topographically complex, or experiencing disturbance. As a result, there is a well-established interest in sensor-based approaches that can facilitate large-scale ecological monitoring while reducing observer bias and disturbance (Bush et al., 2017; Tosa et al., 2021).

Among these approaches, passive acoustic monitoring (PAM) has emerged as a versatile and widely applicable tool for wildlife research and biodiversity assessment (Darras et al., 2025). PAM uses autonomous recording units to collect sound from the environment continuously over long periods, capturing spontaneous vocalizations and other sonotypes from birds, mammals, amphibians, insects, and other taxa, along with anthropogenic and abiotic sounds (Gibb et al., 2019; Sugai et al., 2019).

* Corresponding author at: USDA Forest Service, Pacific Northwest Research Station, Corvallis, OR 97331, USA.
*E-mail address:* damon.lesmeister@usda.gov (D.B. Lesmeister).

https://doi.org/10.1016/j.ecoinf.2026.103657
Received 29 September 2025; Received in revised form 9 February 2026; Accepted 9 February 2026
Available online 11 February 2026

Because many species produce distinctive acoustic signals, PAM enables non-invasive monitoring of diverse taxa via the identification of their acoustic signature in audio recordings (Blumstein et al., 2011; Darras et al., 2018). Applications range from estimating species presence and vocal activity (Campos-Cerqueira and Aide, 2016), to mapping habitat use (Duarte et al., 2024; Rugg et al., 2023), evaluating the impacts of ecosystem change (Ross et al., 2023), and characterizing entire acoustic communities (Burivalova et al., 2019; Farina et al., 2024). PAM is particularly valuable for taxa such as forest birds (Darras et al., 2019; Duchac et al., 2021; Gaylord et al., 2023), marine mammals (Van Parijs et al., 2009), and conservation-relevant or invasive species in complex landscapes (Bota et al., 2024; Duarte et al., 2024; Duchac et al., 2020).

The growing scale of PAM deployments, which routinely generate thousands to millions of hours of audio every year, has outpaced the capacity of manual review for detecting and classifying sounds, making automated methods a central focus of methodological development (Stowell, 2022). In recent years, deep learning applications, particularly convolutional neural networks, have transformed acoustic data processing by offering accurate and efficient detection of sonotypes, most commonly by applying computer vision models to spectrograms (Kahl et al., 2021). These models can learn complex features directly from annotated training data, have the ability to detect many taxa simultaneously, and often generalize well to new acoustic environments, outperforming traditional recognizers based on manually engineered features (Knight et al., 2017; Priyadarshani et al., 2018). Open datasets, benchmark challenges, and simulated soundscapes (Morfi et al., 2021; Stowell et al., 2019; Weldy et al., 2025) have further accelerated progress, while software packages and cloud-based tools have reduced barriers to entry for researchers and practitioners (Appel et al., 2025; Kahl et al., 2021; Lapp et al., 2023; Sethi et al., 2020). Still, many deep learning applications remain limited by model performance, taxonomic breadth, geographic specificity, or integration into real-world ecological workflows.

PNW-Cnet is a convolutional neural network that we developed to enable broad-scale multi-species wildlife detection in passive acoustic recordings from the forests of the Pacific Northwest, USA. Having initially developed PNW-Cnet to identify vocalizations of six forest-adapted owls (Ruff et al., 2020), we have iteratively expanded this model through successive retraining on human-verified detections, and the model now includes 135 sonotypes (hereafter, classes) encompassing 84 species of birds ($n = 72$), mammals ($n = 11$), and amphibians (n = 1), along with key anthropogenic and environmental noise sources (Table 1). This progression reflects a deliberate shift from species-specific monitoring toward community-level biodiversity assessment. The model is operationalized through a data processing pipeline and multiple user-friendly tools, including a Shiny-based desktop application and a Python package (pycnet-audio), that allow flexible, scalable processing of PAM datasets by end users (Ruff et al., 2021; Ruff et al., 2023). Moreover, because we have trained PNW-Cnet on manually annotated field recordings, the model is directly applicable to processing similar recordings, thus avoiding problems of domain shift. We have combined this approach with a targeted data augmentation strategy to enable the inclusion of underrepresented classes, allowing the model to reliably detect rare species and complex environmental sounds. In this study, we describe the development and performance of PNW-Cnet v5, its integration into PAM workflows, and its application to landscape-scale biodiversity monitoring.

## 2. Materials and methods

### 2.1. Study area and monitoring context

PNW-Cnet was developed to support PAM under the Northwest Forest Plan (NWFP), which governs approximately 10 million hectares of federally managed forests across California, Oregon, and Washington, USA (Fig. 1). The NWFP was implemented in response to declining populations of the federally threatened northern spotted owl (*Strix occidentalis caurina*). In contrast to historical federal forest management practices focused primarily on timber harvest, the NWFP emphasized biodiversity and ecosystem health and placed particular importance on maintaining old-growth-associated species, and northern spotted owl populations continue to be monitored under the NWFP into the present day (Lesmeister et al., 2018; Spies et al., 2019).

Each field season, over 4000 autonomous recording units are deployed across the NWFP area using a stratified random sampling design within a hexagonal grid overlaid on federal lands, recording more than two million hours of audio annually (Lesmeister et al., 2021; Lesmeister and Jenkins, 2022). These recordings reflect a diverse acoustic sample from a broad range of forest types, elevations, and ecological conditions. While the original purpose of this monitoring framework was to detect spotted owls and competing barred owls (*Strix varia*), the resulting dataset encodes information on a broad suite of forest wildlife and environmental conditions. The program's focus has thus expanded to include analyses of occupancy, habitat use, species interactions, and disturbance impacts across broad spatial and temporal scales for many vocal species. As data volumes and interest in biodiversity monitoring have grown, automated detection using deep learning applications has become essential for timely processing, model-based inference, and broader integration into forest management planning.

### 2.2. Dataset and target classes

We trained PNW-Cnet v5 using a dataset of 824,120 manually labeled spectrograms derived primarily from field recordings collected by the NWFP monitoring program between 2018 and 2022 (Table 1). All audio was originally recorded at 32 kHz and downsampled to 8 kHz during spectrogram generation. We generated spectrograms using SoX (version 14.4) with a Hann window size of 2048, 50% overlap, and a discrete Fourier transform size of 256. Each spectrogram represents a 12-s segment of audio in the frequency range 0–4000 Hz and was saved as an 8-bit grayscale portable network graphic (PNG) file at a resolution of $1000 \times 257$ pixels. These parameters were selected to capture the fundamental frequency range of most target species and to remain consistent with prior versions of the model.

The training dataset was manually labeled with 135 distinct target classes. These included vocalizations of 84 vertebrate species, primarily birds and mammals, as well as common anthropogenic and environmental sounds such as chainsaws, vehicles, human speech, rain, and wind (Table 1). Class definitions were guided by acoustic distinctiveness, ecological importance, and the need to disambiguate common sources of misclassification. Some classes represented individual species, while others grouped acoustically similar taxa (e.g., chickadee [*Poecile* spp.] calls) or distinct vocalization types within a species (e.g., territorial versus inspection calls for *Strix varia*). Compared to the previous version (PNW-Cnet v4; Ruff et al., 2023), PNW-Cnet v5 added 84 new classes, including 50 classes encompassing 45 new wildlife taxa, 20 classes representing additional call types for taxa previously included in PNW-Cnet v4, and 14 new classes representing anthropogenic and environmental sounds such as aircraft, trains, and thunder (Table 1).

We included a minimum of 500 labeled examples of each class in the training set to improve training set balance, with labels manually verified by expert reviewers. In the case of several rare classes, our audio data contained <500 unique annotated examples (and < 50 examples in several extreme cases, e.g. Cooper's hawk *Astur cooperii*). We employed training data augmentation strategies to ensure adequate model performance for these classes; see “Data augmentation and class balancing” below for details. In total, the training dataset reflects a highly diverse acoustic environment and incorporates a great deal of natural variability in background noise and signal-to-noise ratio, supports multi-species detection and minimizing confusion among frequently co-occurring sounds and ensures the resulting model can be deployed effectively in landscapes across much of western North America.

**Table 1**
Number of images used to train PNW-Cnet v5 by target class. The full set of training, validation, and testing data contained 824,120 images which collectively had 1,148,765 class labels. Training images were used to update model weights during and following each epoch. Validation images were used to evaluate overall model performance after each epoch using the binary cross-entropy loss function; model weights were saved only after epochs in which validation loss decreased. Testing images were not used in training but were used to evaluate the class-specific performance of fully trained candidate models after each completed training run. "Pct. Synth" indicates the percentage of training / validation / testing images that were created using data augmentation. Classes that were newly added in PNW-Cnet v5 are marked with an asterisk (*).

| Class | | Sound | Scientific Name | Training | Validation | Testing | Total | Pct. Synth |
|---|---|---|---|---|---|---|---|---|
| 1 | | Northern saw-whet owl series call | *Aegolius acadicus* | 10,772 | 1264 | 595 | 12,631 | 2.1 |
| 2 | * | Northern saw-whet owl hoots, wails | | 441 | 50 | 10 | 501 | 78.0 |
| 3 | * | Long-eared owl hoot | *Asio otus* | 459 | 61 | 29 | 549 | 66.7 |
| 4 | | Great horned owl hoot | *Bubo virginianus* | 13,249 | 1535 | 781 | 15,565 | 2.6 |
| 5 | * | Great horned owl misc. calls | | 4266 | 534 | 236 | 5036 | 29.0 |
| 6 | | Northern pygmy-owl hoot series | *Glaucidium califonicum* | 20,816 | 2422 | 1179 | 24,417 | 11.8 |
| 7 | | Western screech-owl trill | *Megascops kennicottii* | 14,075 | 1586 | 837 | 16,498 | 4.2 |
| 8 | * | Western screech-owl bark | | 1977 | 210 | 138 | 2325 | 68.1 |
| 9 | * | Western screech-owl juvenile | | 4357 | 531 | 254 | 5142 | 57.3 |
| 10 | | Flammulated owl hoot series | *Psiloscops flammeolus* | 17,967 | 2245 | 1052 | 21,264 | 1.6 |
| 11 | * | Great grey owl hoot series | *Strix nebulosa* | 2944 | 334 | 139 | 3417 | 66.8 |
| 12 | * | Great grey owl misc. calls | | 560 | 74 | 32 | 666 | 66.8 |
| 13 | | Spotted owl four-note location call | *Strix occidentalis caurina* | 32,522 | 3929 | 1905 | 38,356 | 2.7 |
| 14 | | Spotted owl series call | | 10,954 | 1208 | 632 | 12,794 | 2.1 |
| 15 | * | Strix owl bark | *Strix* spp. | 4313 | 486 | 227 | 5026 | 36.3 |
| 16 | | Strix owl contact whistle | | 5012 | 598 | 306 | 5916 | 1.6 |
| 17 | | Barred owl eight-note call | *Strix varia* | 21,515 | 2590 | 1245 | 25,350 | 1.4 |
| 18 | | Barred owl inspection call | | 15,166 | 1801 | 906 | 17,873 | 2.0 |
| 19 | | Barred owl series call | | 14,811 | 1737 | 887 | 17,435 | 1.0 |
| 20 | * | Sandhill crane | *Antigone canadensis* | 560 | 59 | 42 | 661 | 66.9 |
| 21 | * | Ruffed grouse drum | *Bonasa umbellus* | 1027 | 135 | 60 | 1222 | 68.3 |
| 22 | | Marbled murrelet keer call | *Brachyramphus marmoratus* | 20,178 | 2391 | 1188 | 23,757 | 5.5 |
| 23 | * | Marbled murrelet groan | | 874 | 89 | 56 | 1019 | 67.1 |
| 24 | | Canada goose honk | *Branta canadensis* | 6053 | 694 | 400 | 7147 | 5.9 |
| 25 | * | California quail 3-note song | *Callipepla californica* | 2107 | 244 | 117 | 2468 | 67.4 |
| 26 | | Common nighthawk peent | *Chordeiles minor* | 11,433 | 1318 | 686 | 13,437 | 2.6 |
| 27 | | Common nighthawk boom | | 8685 | 1006 | 495 | 10,186 | 0.9 |
| 28 | | Sooty grouse drum | *Dendragapus fuliginosus* | 13,216 | 1594 | 817 | 15,627 | 7.0 |
| 29 | * | Sooty grouse series call | | 897 | 102 | 56 | 1055 | 67.0 |
| 30 | * | Wilson's snipe display | *Gallinago delicata* | 719 | 93 | 52 | 864 | 69.1 |
| 31 | * | Wild turkey gobble | *Meleagris gallapavo* | 2543 | 287 | 134 | 2964 | 69.3 |
| 32 | | Mountain quail single call | *Oreortyx pictus* | 10,040 | 1157 | 584 | 11,781 | 15.3 |
| 33 | * | Mountain quail clucks, whinnies | | 983 | 124 | 70 | 1177 | 69.0 |
| 34 | | Band-tailed pigeon song | *Patagioenas fasciata* | 11,393 | 1299 | 669 | 13,361 | 6.5 |
| 35 | * | Band-tailed pigeon call | | 1357 | 151 | 74 | 1582 | 67.3 |
| 36 | | Common poorwill | *Phalaenoptilus nuttallii* | 8593 | 1042 | 477 | 10,112 | 1.8 |
| 37 | * | Eurasian collared dove | *Streptopelia decaocto* | 877 | 94 | 52 | 1023 | 67.5 |
| 38 | | Mourning dove | *Zenaida macrourus* | 5941 | 753 | 338 | 7032 | 7.6 |
| 39 | * | American crow | *Corvus brachyrhyncos* | 2669 | 345 | 155 | 3169 | 67.5 |
| 40 | | Common raven | *Corvus corax* | 21,523 | 2579 | 1272 | 25,374 | 8.5 |
| 41 | | Steller's jay series call | *Cyanocitta stelleri* | 23,549 | 2789 | 1331 | 27,669 | 13.0 |
| 42 | * | Steller's jay rattle (non-diagnostic) | | 462 | 56 | 18 | 536 | 75.0 |
| 43 | | Clark's nutcracker | *Nucifraga columbiana* | 4390 | 478 | 242 | 5110 | 44.5 |
| 44 | | Canada jay | *Perisoreus canadensis* | 14,705 | 1724 | 943 | 17,372 | 4.6 |
| 45 | | Northern flicker series call | *Colaptes auratus* | 13,713 | 1642 | 833 | 16,188 | 10.6 |
| 46 | | Northern flicker skew | | 4539 | 522 | 264 | 5325 | 54.6 |
| 47 | | Downy woodpecker | *Dryobates pubescens* | 4299 | 501 | 275 | 5075 | 49.4 |
| 48 | | Pileated woodpecker cackle / wuk | *Dryocopus pileatus* | 10,310 | 1203 | 634 | 12,147 | 6.5 |
| 49 | * | White-headed woodpecker call | *Leuconotopicus albolarvatus* | 4441 | 475 | 277 | 5193 | 64.7 |
| 50 | * | Hairy woodpecker rattle | *Leuconotopicus villosus* | 2693 | 302 | 135 | 3130 | 68.4 |
| 51 | * | Hairy woodpecker misc. calls | | 451 | 55 | 19 | 525 | 82.1 |
| 52 | * | Acorn woodpecker | *Melanerpes formicivorus* | 1950 | 227 | 112 | 2289 | 68.4 |
| 53 | | Sapsucker spp. drum | *Sphyrapicus* spp. | 4551 | 527 | 268 | 5346 | 34.1 |
| 54 | * | Sapsucker spp. call | | 466 | 58 | 37 | 561 | 82.2 |
| 55 | * | Williamson's sapsucker drum | *Sphyrapicus thyroideus* | 435 | 45 | 32 | 512 | 79.1 |
| 56 | | Woodpecker (non-sapsucker) drum | | 12,044 | 1398 | 700 | 14,142 | 13.9 |
| 57 | * | Wilson's warbler song | *Cardellina pusilla* | 478 | 61 | 20 | 559 | 75.0 |
| 58 | | Hermit thrush song | *Catharus guttatus* | 25,093 | 2951 | 1413 | 29,457 | 16.5 |
| 59 | * | Hermit thrush chit call | | 515 | 66 | 40 | 621 | 69.4 |
| 60 | * | Hermit thrush scratchy call | | 4591 | 564 | 283 | 5438 | 39.5 |
| 61 | | Swainson's thrush song | *Catharus ustulatus* | 16,166 | 1919 | 944 | 19,029 | 13.4 |
| 62 | * | Swainson's thrush call | | 5403 | 615 | 330 | 6348 | 5.2 |
| 63 | | Wrentit song | *Chamaea fasciata* | 7219 | 868 | 421 | 8508 | 12.3 |
| 64 | | Olive-sided flycatcher song | *Contopus cooperi* | 12,428 | 1431 | 778 | 14,637 | 13.1 |
| 65 | * | Olive-sided flycatcher call | | 4606 | 561 | 286 | 5453 | 26.7 |
| 66 | * | Western wood-pewee | *Contopus sordidulus* | 4713 | 559 | 261 | 5533 | 67.2 |
| 67 | * | Western flycatcher | *Empidonax difficilis* | 2182 | 271 | 135 | 2588 | 68.9 |
| 68 | * | Dusky flycatcher call | *Empidonax oberholseri* | 5059 | 602 | 279 | 5940 | 60.8 |
| 69 | * | Purple finch | *Haemorhous purpureus* | 1470 | 193 | 85 | 1748 | 68.6 |

(*continued on next page*)

**Table 1** (*continued*)

| Class | | Sound | Scientific Name | Training | Validation | Testing | Total | Pct. Synth |
|---|---|---|---|---|---|---|---|---|
| 70 | * | Evening grosbeak call | *Hesperiphona verspertina* | 582 | 66 | 27 | 675 | 70.2 |
| 71 | | Varied thrush song | *Ixoreus naevius* | 29,111 | 3357 | 1719 | 34,187 | 11.3 |
| 72 | * | Varied thrush call | | 2479 | 292 | 146 | 2917 | 67.2 |
| 73 | * | Dark-eyed junco | *Junco hyemalis* | 4447 | 547 | 289 | 5283 | 54.4 |
| 74 | * | Orange-crowned warbler | *Leiothlypis celata* | 1214 | 143 | 69 | 1426 | 70.3 |
| 75 | * | Red crossbill call | *Loxia curvirostra* | 873 | 97 | 58 | 1028 | 68.1 |
| 76 | | Townsend's solitaire call | *Myadestes townsendi* | 12,777 | 1474 | 718 | 14,969 | 7.8 |
| 77 | * | Black-headed grosbeak | *Pheucticus melanocephalus* | 2714 | 333 | 156 | 3203 | 70.2 |
| 78 | * | Spotted towhee song | *Pipilo maculatus* | 2825 | 344 | 176 | 3345 | 73.7 |
| 79 | | Spotted towhee call | | 5528 | 673 | 351 | 6552 | 13.0 |
| 80 | * | Western tanager song | *Piranga ludoviciana* | 3972 | 477 | 230 | 4679 | 72.2 |
| 81 | * | Western tanager call | | 3173 | 343 | 176 | 3692 | 72.9 |
| 82 | | Chickadee song | *Poecile* spp. | 7118 | 846 | 416 | 8380 | 26.9 |
| 83 | * | Chickadee call | | 4477 | 529 | 257 | 5263 | 73.6 |
| 84 | * | Mountain bluebird | *Sialia currucoides* | 477 | 71 | 24 | 572 | 67.1 |
| 85 | | Nuthatch spp. song | *Sitta* spp. | 29,102 | 3502 | 1680 | 34,284 | 22.1 |
| 86 | * | Nuthatch spp. call | | 954 | 121 | 61 | 1136 | 69.5 |
| 87 | * | Pine siskin call | *Spinus pinus* | 1487 | 187 | 82 | 1756 | 70.1 |
| 88 | * | Chipping sparrow | *Spizella passerina* | 484 | 54 | 29 | 567 | 82.4 |
| 89 | * | House wren | *Troglodytes aedon* | 458 | 52 | 25 | 535 | 84.3 |
| 90 | | American robin whinny | *Turdus migratorius* | 10,023 | 1162 | 583 | 11,768 | 9.9 |
| 91 | * | American robin song and misc. calls | | 8490 | 989 | 488 | 9967 | 16.9 |
| 92 | * | Hutton's vireo song / call | *Vireo huttoni* | 989 | 114 | 57 | 1160 | 69.0 |
| 93 | * | White-crowned sparrow | *Zonotrichia leucophrys* | 441 | 57 | 31 | 529 | 81.3 |
| 94 | * | Sharp-shinned hawk series call | *Accipiter striatus* | 468 | 53 | 26 | 547 | 66.7 |
| 95 | * | Raptor scream call | Accipitrid | 2464 | 286 | 183 | 2933 | 67.4 |
| 96 | * | American goshawk series call | *Astur atricapillus* | 4986 | 626 | 273 | 5885 | 69.4 |
| 97 | * | American goshawk scream | | 5132 | 583 | 284 | 5999 | 50.5 |
| 98 | * | Cooper's hawk series call | *Astur cooperii* | 432 | 50 | 18 | 500 | 95.8 |
| 99 | * | Red-tailed hawk scream | *Buteo jamaicensis* | 1007 | 125 | 64 | 1196 | 70.3 |
| 100 | * | Red-tailed hawk misc. call | | 470 | 50 | 25 | 545 | 84.0 |
| 101 | * | Merlin | *Falco columbarius* | 437 | 36 | 27 | 500 | 75.0 |
| 102 | * | American kestrel | *Falco sparverius* | 416 | 55 | 31 | 502 | 80.3 |
| 103 | * | Bald eagle | *Haliaeetus leucocephalus* | 428 | 60 | 24 | 512 | 76.8 |
| 104 | * | Osprey | *Pandion haliaetus* | 1056 | 117 | 67 | 1240 | 67.2 |
| 105 | * | Coyote | *Canis latrans* | 1083 | 128 | 70 | 1281 | 66.9 |
| 106 | | Wolf howl | *Canis lupus* | 1290 | 118 | 92 | 1500 | 66.7 |
| 107 | * | Elk | *Cervus canadensis* | 652 | 68 | 28 | 748 | 67.4 |
| 108 | | Chipmunk spp. chirp | *Neotamias* spp. | 18,574 | 2138 | 1075 | 21,787 | 9.8 |
| 109 | | American pika | *Ochotona princeps* | 4364 | 545 | 275 | 5184 | 8.7 |
| 110 | * | Deer snort | *Odocoileus* spp. | 440 | 47 | 20 | 507 | 81.1 |
| 111 | | Douglas squirrel rattle | *Tamiasciurus douglasii* | 13,182 | 1578 | 767 | 15,527 | 10.9 |
| 112 | | Douglas squirrel chirp | | 12,101 | 1474 | 726 | 14,301 | 9.7 |
| 113 | * | Black bear juvenile | *Ursus americanus* | 398 | 65 | 37 | 500 | 94.8 |
| 114 | * | Wildcat scream | | 416 | 59 | 25 | 500 | 95.8 |
| 115 | * | Airplane | | 7675 | 931 | 453 | 9059 | 6.9 |
| 116 | * | Chainsaw | | 4808 | 540 | 242 | 5590 | 5.7 |
| 117 | * | Military jet | | 961 | 108 | 57 | 1126 | 0.0 |
| 118 | | Gunshot | | 5270 | 601 | 318 | 6189 | 1.6 |
| 119 | * | Highway noise | | 4475 | 514 | 253 | 5242 | 9.4 |
| 120 | * | Car horn | | 419 | 52 | 32 | 503 | 69.4 |
| 121 | | Human speech | *Homo sapiens* | 7553 | 901 | 490 | 8944 | 1.5 |
| 122 | * | Spotted owl survey tone | | 8489 | 1046 | 501 | 10,036 | 99.6 |
| 123 | * | Train | | 2267 | 273 | 138 | 2678 | 66.9 |
| 124 | | Yarder (machine) | | 6043 | 737 | 345 | 7125 | 4.5 |
| 125 | * | Bullfrog | *Rana catesbeiana* | 1545 | 176 | 82 | 1803 | 67.1 |
| 126 | | Frog chorus | | 12,796 | 1478 | 746 | 15,020 | 4.5 |
| 127 | | Buzzing insect | | 31,764 | 3700 | 1968 | 37,432 | 13.1 |
| 128 | * | Chicken | *Gallus gallus domesticus* | 4417 | 490 | 220 | 5127 | 62.2 |
| 129 | * | Cow | *Bos taurus* | 3814 | 454 | 221 | 4489 | 67.9 |
| 130 | * | Stream noise | | 8218 | 1007 | 465 | 9690 | 0.9 |
| 131 | | Dog | *Canis familiaris* | 19,749 | 2284 | 1078 | 23,111 | 2.6 |
| 132 | * | Rain | | 63,516 | 7427 | 3700 | 74,643 | 54.2 |
| 133 | * | Thunder | | 1172 | 119 | 73 | 1364 | 66.9 |
| 134 | * | Tree creaking | | 13,694 | 1602 | 816 | 16,112 | 0.6 |
| 135 | * | Cricket | | 43,187 | 4982 | 2488 | 50,657 | 0.4 |

### *2.3. Model architecture and training*

PNW-Cnet is a deep convolutional neural network implemented in Python using the Keras API with a TensorFlow backend (Abadi et al., 2015; Watson et al., 2024). PNW-Cnet accepts grayscale spectrograms as input and detects distinctive visual patterns corresponding to target sonotypes. For each input spectrogram, PNW-Cnet produces a vector of 135 class scores representing the model's confidence (i.e., confidence score) that the corresponding 12-s audio segment contains an audio signature matching each of the 135 target classes. Because each spectrogram may contain multiple sonotypes, we implemented the model as a multi-label classifier using independent sigmoid activations for each class. This enabled the network to assign high scores to multiple classes per input, reflecting real-world acoustic overlap and supporting flexible

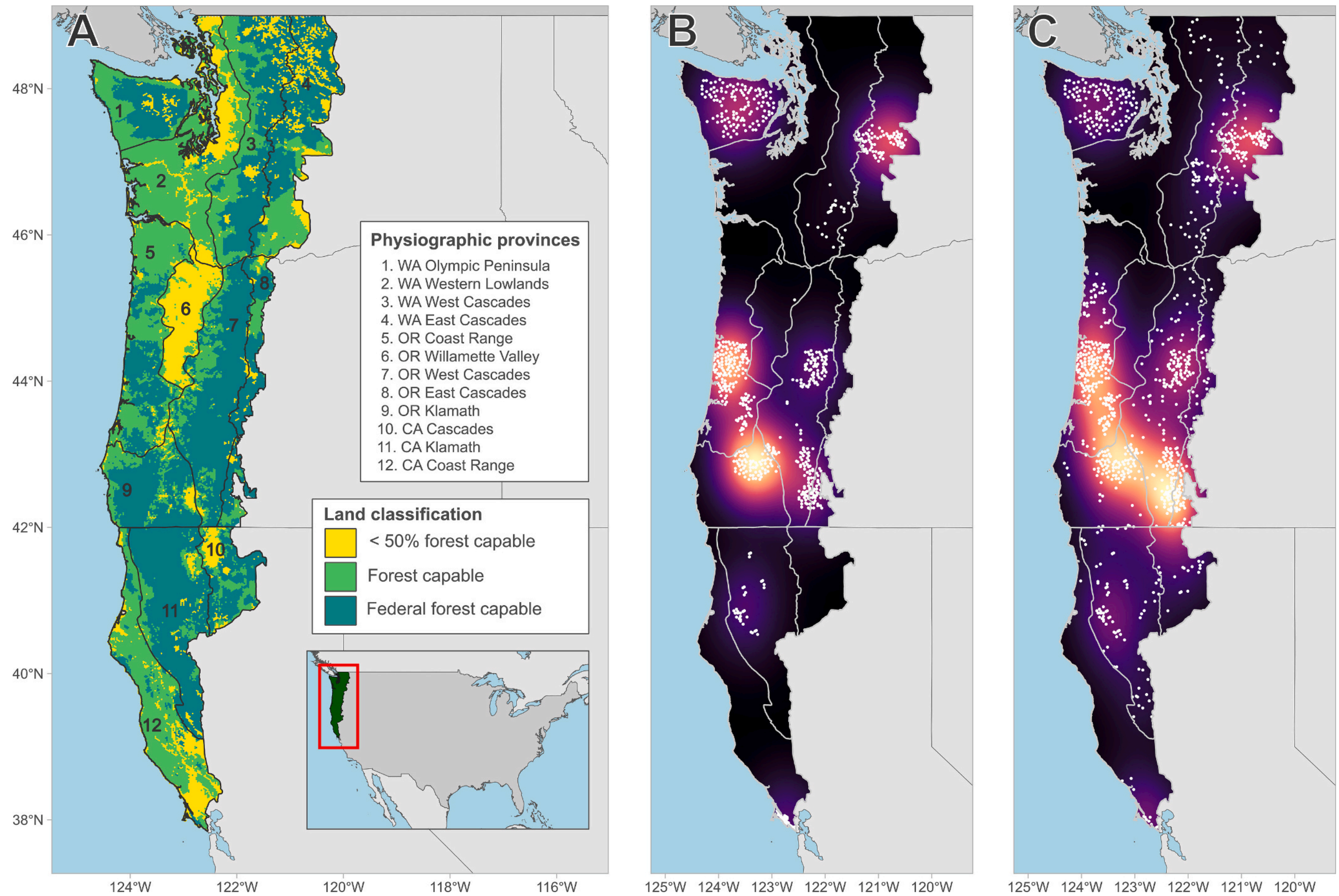


**Fig. 1.** Panel A: Geographic area of the Northwest Forest Plan (NWFP) passive acoustic monitoring program in California (CA), Oregon (OR), and Washington (WA), USA. NWFP monitoring sites were selected from a layer of 5-km$^2$ hexagons covering the geographic range of the northern spotted owl. Only hexagons that were ≥ 50% "forest capable" (i.e., capable of supporting forest) and ≥ 25% federally owned were eligible for sampling. Inset figure shows the location of the NWFP monitoring area within the contiguous United States of America. The NWFP area is divided into 12 physiographic provinces intended to reflect landscape-scale differences in climate and forest type. Panel B: Geographic origins of training data for PNW-Cnet v5. Most of the training data (n = ~655,000 images, 80% of total) were unmodified examples of target classes drawn from data recorded at 701 long-term acoustic monitoring hexagons (white dots) in California, Oregon, and Washington between 2017 and 2022. The number of such examples contributed by each field site was projected onto a raster with resolution of 0.01 decimal degrees, which was then smoothed with the terra:focalMat function in R using a Gaussian weight matrix with σ = 0.35. The remaining training data (20% of total) included modified or synthesized examples or from alternate study areas and were excluded from this visualization. Panel C: Geographic origins of an independent test set of 8500 images used to estimate PNW-Cnet v5 precision and determine optimal detection threshold for 85 target classes. Test data were drawn from recordings collected at 968 field sites (white dots) in California, Oregon, and Washington in 2023. Images contributed by each site ranged from 1 to 110 (mean = 8.78, standard deviation = 9.50).

downstream thresholding.

The network architecture included six convolutional layers and two fully connected layers. Convolutional layers used 5 × 5 filters with ReLU activation and were each followed by 2 × 2 max pooling. Filter counts increased with depth: 32, 32, 64, 64, 128, and 128. The final convolutional output was flattened and passed through a fully connected layer with 512 nodes and ReLU activation. The output layer consisted of 135 sigmoid-activated nodes corresponding to the target classes. During training, each convolutional layer was followed by 30% dropout, as was the 512-node fully connected layer.

Training was performed for 30 epochs with a batch size of 256 using the Adam optimizer with an initial learning rate of 0.0015 and exponential decay of 0.96 every 10,000 batches. The loss function was binary cross-entropy, and the training procedure also reported accuracy ([true positives + true negatives] / [true positives + false positives + true negatives + false negatives]), aggregated across all classes on the training and validation sets at the end of each epoch. To reduce overfitting, early stopping was implemented by saving model weights only when validation loss decreased. All training was conducted on an IBM PPC64LE machine with four Nvidia V100 GPUs (16 GB each). Training configuration scripts, spectrogram generation code, and version-controlled training pipelines are archived in an online repository to ensure reproducibility; see "Data Availability" below.

### *2.4. Data augmentation and class balancing*

To address class imbalance in the training set and enable the inclusion of rare classes, we implemented a targeted augmentation procedure for underrepresented classes. For any class with fewer than 5000 unique examples, we generated synthetic spectrograms by overlaying verified vocalizations onto background noise clips drawn from a library of "clean" 12-s segments. We selected these background clips from previous years' field recordings that had been screened to ensure that no target classes were present in the images.

Each augmented audio clip preserved the spectral and temporal characteristics of the target class while introducing variation in background acoustic context, simulating real-world variability in signal-to-noise ratio, ambient noise, and call clarity. Augmentation was performed until each rare class reached either 5000 images or three times the original number of examples, with a minimum floor of 500 images per class. All generated spectrograms were visually inspected to ensure clarity and class integrity. These augmented images were used alongside those generated directly from field recordings during model training to improve robustness and reduce overfitting to dominant classes.

### 2.5. Model evaluation

We evaluated model performance using two independent datasets: a held-out test set and a manually reviewed field-based evaluation set. The held-out test set consisted of 41,202 labeled spectrograms which we selected randomly from the training set and withheld from training. For each class, we calculated precision (true positives / [true positives + false positives]) and recall (true positives / [true positives + false negatives]) at detection thresholds of 0.25, 0.50, 0.75, and 0.95. These thresholds reflect common trade-offs between broad detection and high-confidence classification and were used to construct class-level performance profiles. Therefore, we calculated precision and recall at these thresholds for all classes. We also calculated the F1 score ([2 * Precision * Recall] / [Precision + Recall], used to find an "optimal" threshold if precision and recall are weighted equally) for all classes at thresholds from 0.01 to 0.99 and determined the threshold at which F1 score was highest for each class (Table 2).

To assess generalization and guide threshold selection in practical applications, we conducted an independent evaluation using field data collected from the 2023 PAM deployment within the study area. We selected classes with ≥100 apparent detections at and above a 0.80 confidence score threshold after first filtering them based on species' known geographic ranges and seasonal presence using expert knowledge and eBird range maps (ebird.org). For each class, we sampled 100 detections, stratified to ensure proportional representation of different physiographic provinces comprising the NWFP monitoring area (Fig. 1). We manually reviewed the resulting set of clips, organized by class, using Kaleidoscope Pro (Wildlife Acoustics, Maynard, MA, USA). Based on spectrogram and audio playback, each detection was marked as a true positive if the target class was present and clearly identifiable or a false positive otherwise. All detections were fully reviewed by two expert reviewers, followed by partial review by a third expert reviewer to resolve disagreements. This review yielded empirical estimates of precision across fine-scale confidence score thresholds (0.800–1.00), enabling identification of optimal thresholds for specific classes and use cases. Precision-recall trade-offs were used to inform post-processing thresholds and to guide threshold calibration and screening filters in field deployments.

### 2.6. Software tools and availability

We developed two companion tools to support and facilitate the practical application of PNW-Cnet v5 in both research and management settings. First, Shiny_PNW-Cnet is a graphical desktop application built in R/Shiny that allows users to process locally stored audio data, visualize detections, and manually review outputs through an accessible interface (https://github.com/zjruff/Shiny_PNW-Cnet). Second, pycnet-audio is a Python package that provides command-line tools and an API for scripting custom workflows and scaling analyses across large acoustic datasets (https://github.com/zjruff/pycnet-audio). Both tools interface directly with model output and offer utilities for detection filtering, thresholding, and quality control.

## 3. Results

### 3.1. Model training and selection

Training of the final selected model was completed using four Nvidia V100 GPUs. Although the model was trained for 30 epochs, the optimal performance was observed at epoch nine, at which point the model achieved a training accuracy of 0.6928 and a training loss of 0.0143. Validation accuracy and loss at this epoch were 0.7084 and 0.0132, respectively. These values were comparable across subsequent epochs, but model weights from epoch nine were retained as validation loss did not decrease in subsequent epochs. Selection of this trained model was guided by performance on priority species of management interest, particularly northern spotted owl, barred owl, and marbled murrelet (*Brachyramphus marmoratus*), as well as consistency across classes and avoidance of overfitting.

### 3.2. Performance on the held-out test set

We evaluated the model using a held-out test set of 41,202 labeled spectrograms and found that average precision across all classes was high at every threshold, ranging from 0.818 at 0.25 to 0.982 at 0.95 (Table 2). As expected, precision improved as thresholds increased, reflecting reduced rates of false positives. However, this gain came at the cost of recall, which declined from 0.590 at threshold 0.25 to 0.335 at threshold 0.95. The relationship between threshold and recall was nonlinear, with more substantial losses occurring at higher cutoffs (Fig. 2).

At the most conservative threshold (0.95), 23 classes produced no apparent detections and were excluded from mean precision calculations; these included some of the newly added wildlife taxa and rarer classes, which were under-represented in the training set relative to already established classes. For recall metrics, these classes were assigned a value of zero. Only four classes failed to produce any detections at the lowest threshold (0.25), demonstrating that the model is capable of low-confidence detection for nearly all classes under permissive settings.

Performance varied notably by class and training history. The 51 legacy classes inherited from PNW-Cnet v4 exhibited stronger performance overall, with average precision ranging from 0.871 to 0.987 across thresholds. Recall was also higher and more stable across thresholds (0.799 at 0.25 threshold to 0.536 at 0.95 threshold), suggesting benefits from iterative training, greater sample depth, and improved representation in the training set. In contrast, some classes, particularly those based on limited field examples and trained with a higher proportion of synthetically augmented data (Table 1), exhibited steep declines in detections as thresholds increased (Fig. 2). Maximum F1 score varied from 0.085 to 0.995, and maximum F1 score was achieved at thresholds ranging from 0.02 to 0.71, depending on class (Table 2).

Performance also varied by class (Fig. 2). For example, classes representing highly-stereotyped vocalizations of common forest birds (e.g., Pileated woodpecker *Dryocopus pileatus*, Olive-sided flycatcher *Contopus cooperi*) and distinctive nocturnal vocalizers (e.g., spotted owl, barred owl) performed well, while those representing more variable vocalizations (e.g. Western tanager *Piranga ludoviciana*, Black-headed grosbeak *Pheucticus melanocephalus*, Coyote *Canis latrans*) showed lower performance (Fig. 2). Several anthropogenic classes such as human speech, chainsaw, and vehicle also exhibited high precision and moderate recall, indicating that the model is well-suited to identifying key sources of environmental noise.

### 3.3. Independent validation with field data

In validation of the second test dataset, which was comprised of randomly selected apparent detections from field recordings, 77 classes achieved precision ≥0.95 at thresholds between 0.80 and 0.975 (Fig. 3). Among these, 54 classes achieved precision ≥0.95 even at the lowest threshold of 0.80, indicating that performance was strong for many classes even at modest thresholds. Some classes required higher confidence score thresholds to achieve precision ≥0.95, and eight classes did not achieve precision ≥0.95 at any threshold, with the maximum achieved precision for these classes ranging from 0 (i.e., no true positives) to 0.89 (Fig. 3). The number of detections declined at higher confidence score thresholds for all classes, but the rate of decrease varied by class

**Table 2**

Class-specific performance of PNW-Cnet v5 at several score thresholds (0.25, 0.50, 0.75, 0.95) based on a test set containing 41,202 images that were held out from the full training dataset. “In test set” refers to the number of images in the test set with each label. At each threshold, “Apparent detections” refers to the number of images to which PNW-Cnet v5 assigned a score greater than or equal to the threshold; “Precision” refers to the proportion of apparent detections that were true positives; and “Recall” refers to the proportion of real examples of each class in the test set that were assigned a score greater than or equal to the threshold, i.e., the proportion of positive examples that were successfully detected by the model. “Max F1” is the highest F1 score observed for the test set at any threshold, calculated as (2 * Precision * Recall) / (Precision + Recall) and “At threshold” refers to the threshold at which the highest F1 score was achieved. See Table 1 for description of class number.

| Class | In test set | Apparent detections | | | | Precision | | | | Recall | | | | Max F1 | At threshold |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| | | (0.25) | (0.50) | (0.75) | (0.95) | (0.25) | (0.50) | (0.75) | (0.95) | (0.25) | (0.50) | (0.75) | (0.95) | | |
| 1 | 595 | 660 | 617 | 593 | 539 | 0.865 | 0.909 | 0.929 | 0.963 | 0.960 | 0.943 | 0.926 | 0.872 | 0.928 | 0.65 |
| 2 | 10 | 2 | 1 | 0 | 0 | 1.000 | 1.000 | NA | NA | 0.200 | 0.100 | 0.000 | 0.000 | 0.500 | 0.13 |
| 3 | 29 | 26 | 17 | 10 | 1 | 0.615 | 0.824 | 1.000 | 1.000 | 0.552 | 0.483 | 0.345 | 0.034 | 0.627 | 0.36 |
| 4 | 781 | 741 | 676 | 623 | 474 | 0.927 | 0.967 | 0.979 | 0.996 | 0.880 | 0.837 | 0.781 | 0.604 | 0.906 | 0.27 |
| 5 | 236 | 212 | 159 | 124 | 75 | 0.722 | 0.849 | 0.887 | 0.987 | 0.648 | 0.572 | 0.466 | 0.314 | 0.706 | 0.30 |
| 6 | 1179 | 1214 | 1134 | 1078 | 955 | 0.880 | 0.919 | 0.936 | 0.956 | 0.906 | 0.884 | 0.856 | 0.774 | 0.902 | 0.49 |
| 7 | 837 | 932 | 825 | 726 | 561 | 0.807 | 0.876 | 0.937 | 0.993 | 0.898 | 0.864 | 0.812 | 0.665 | 0.878 | 0.67 |
| 8 | 138 | 71 | 37 | 18 | 3 | 0.775 | 0.946 | 1.000 | 1.000 | 0.399 | 0.254 | 0.130 | 0.022 | 0.559 | 0.14 |
| 9 | 254 | 280 | 245 | 221 | 174 | 0.864 | 0.935 | 0.964 | 0.989 | 0.953 | 0.902 | 0.839 | 0.677 | 0.926 | 0.33 |
| 10 | 1052 | 1061 | 1023 | 989 | 920 | 0.949 | 0.968 | 0.984 | 0.998 | 0.957 | 0.941 | 0.925 | 0.873 | 0.956 | 0.67 |
| 11 | 139 | 118 | 104 | 96 | 90 | 0.907 | 0.981 | 1.000 | 1.000 | 0.770 | 0.734 | 0.691 | 0.647 | 0.846 | 0.43 |
| 12 | 32 | 10 | 5 | 3 | 1 | 0.800 | 1.000 | 1.000 | 1.000 | 0.250 | 0.156 | 0.094 | 0.031 | 0.400 | 0.28 |
| 13 | 1905 | 2318 | 1746 | 1304 | 700 | 0.741 | 0.873 | 0.937 | 0.981 | 0.902 | 0.800 | 0.641 | 0.361 | 0.838 | 0.38 |
| 14 | 632 | 576 | 413 | 337 | 236 | 0.769 | 0.915 | 0.961 | 0.987 | 0.701 | 0.598 | 0.513 | 0.369 | 0.737 | 0.23 |
| 15 | 227 | 227 | 199 | 178 | 124 | 0.815 | 0.874 | 0.899 | 0.911 | 0.815 | 0.767 | 0.705 | 0.498 | 0.822 | 0.33 |
| 16 | 306 | 264 | 236 | 212 | 157 | 0.898 | 0.958 | 0.976 | 0.987 | 0.775 | 0.739 | 0.676 | 0.507 | 0.840 | 0.28 |
| 17 | 1245 | 1309 | 1120 | 965 | 601 | 0.840 | 0.921 | 0.955 | 0.983 | 0.884 | 0.829 | 0.741 | 0.475 | 0.875 | 0.35 |
| 18 | 906 | 937 | 843 | 753 | 625 | 0.869 | 0.922 | 0.959 | 0.987 | 0.898 | 0.858 | 0.797 | 0.681 | 0.893 | 0.30 |
| 19 | 887 | 694 | 465 | 380 | 214 | 0.814 | 0.959 | 0.995 | 0.995 | 0.637 | 0.503 | 0.426 | 0.240 | 0.718 | 0.23 |
| 20 | 42 | 10 | 8 | 4 | 2 | 0.800 | 0.750 | 1.000 | 1.000 | 0.190 | 0.143 | 0.095 | 0.048 | 0.500 | 0.05 |
| 21 | 60 | 0 | 0 | 0 | 0 | NA | NA | NA | NA | 0.000 | 0.000 | 0.000 | 0.000 | 0.085 | 0.02 |
| 22 | 1188 | 1226 | 1202 | 1166 | 1110 | 0.947 | 0.963 | 0.979 | 0.991 | 0.977 | 0.974 | 0.960 | 0.926 | 0.970 | 0.59 |
| 23 | 56 | 53 | 35 | 19 | 0 | 0.642 | 0.686 | 0.789 | NA | 0.607 | 0.429 | 0.268 | 0.000 | 0.625 | 0.21 |
| 24 | 400 | 305 | 257 | 204 | 122 | 0.866 | 0.899 | 0.961 | 0.992 | 0.660 | 0.578 | 0.490 | 0.302 | 0.749 | 0.23 |
| 25 | 117 | 99 | 88 | 80 | 51 | 0.859 | 0.886 | 0.900 | 0.941 | 0.726 | 0.667 | 0.615 | 0.410 | 0.804 | 0.20 |
| 26 | 686 | 650 | 606 | 573 | 506 | 0.912 | 0.942 | 0.962 | 0.980 | 0.864 | 0.832 | 0.803 | 0.723 | 0.891 | 0.28 |
| 27 | 495 | 511 | 479 | 460 | 424 | 0.914 | 0.958 | 0.974 | 0.986 | 0.943 | 0.927 | 0.905 | 0.844 | 0.946 | 0.53 |
| 28 | 817 | 794 | 706 | 605 | 417 | 0.873 | 0.921 | 0.955 | 0.983 | 0.848 | 0.796 | 0.707 | 0.502 | 0.865 | 0.36 |
| 29 | 56 | 38 | 30 | 21 | 13 | 0.684 | 0.800 | 0.905 | 1.000 | 0.464 | 0.429 | 0.339 | 0.232 | 0.596 | 0.11 |
| 30 | 52 | 29 | 21 | 18 | 9 | 0.897 | 0.952 | 0.944 | 1.000 | 0.500 | 0.385 | 0.327 | 0.173 | 0.740 | 0.10 |
| 31 | 134 | 60 | 31 | 20 | 7 | 0.700 | 0.968 | 1.000 | 1.000 | 0.313 | 0.224 | 0.149 | 0.052 | 0.460 | 0.16 |
| 32 | 584 | 547 | 447 | 352 | 220 | 0.832 | 0.915 | 0.966 | 1.000 | 0.779 | 0.700 | 0.582 | 0.377 | 0.809 | 0.26 |
| 33 | 70 | 50 | 43 | 29 | 20 | 0.880 | 0.907 | 0.966 | 1.000 | 0.629 | 0.557 | 0.400 | 0.286 | 0.733 | 0.25 |
| 34 | 669 | 610 | 583 | 552 | 485 | 0.959 | 0.976 | 0.982 | 0.990 | 0.874 | 0.851 | 0.810 | 0.717 | 0.915 | 0.25 |
| 35 | 74 | 36 | 21 | 18 | 8 | 0.833 | 0.952 | 1.000 | 1.000 | 0.405 | 0.270 | 0.243 | 0.108 | 0.564 | 0.16 |
| 36 | 477 | 446 | 432 | 422 | 404 | 0.982 | 0.986 | 0.988 | 0.993 | 0.918 | 0.893 | 0.874 | 0.841 | 0.952 | 0.18 |
| 37 | 52 | 24 | 13 | 7 | 3 | 0.792 | 1.000 | 1.000 | 1.000 | 0.365 | 0.250 | 0.135 | 0.058 | 0.604 | 0.10 |
| 38 | 338 | 315 | 274 | 233 | 158 | 0.873 | 0.938 | 0.966 | 0.994 | 0.814 | 0.760 | 0.666 | 0.464 | 0.846 | 0.43 |
| 39 | 155 | 125 | 89 | 70 | 36 | 0.712 | 0.809 | 0.871 | 0.972 | 0.574 | 0.465 | 0.394 | 0.226 | 0.654 | 0.30 |
| 40 | 1272 | 1089 | 889 | 735 | 485 | 0.869 | 0.935 | 0.967 | 0.992 | 0.744 | 0.653 | 0.559 | 0.378 | 0.801 | 0.23 |
| 41 | 1331 | 1176 | 906 | 732 | 494 | 0.805 | 0.913 | 0.955 | 0.982 | 0.711 | 0.621 | 0.525 | 0.364 | 0.763 | 0.32 |
| 42 | 18 | 0 | 0 | 0 | 0 | NA | NA | NA | NA | 0.000 | 0.000 | 0.000 | 0.000 | 0.364 | 0.05 |
| 43 | 242 | 222 | 202 | 169 | 115 | 0.914 | 0.955 | 0.982 | 1.000 | 0.839 | 0.798 | 0.686 | 0.475 | 0.886 | 0.13 |
| 44 | 943 | 966 | 882 | 797 | 647 | 0.856 | 0.907 | 0.944 | 0.977 | 0.877 | 0.848 | 0.797 | 0.670 | 0.878 | 0.41 |
| 45 | 833 | 742 | 694 | 628 | 467 | 0.951 | 0.974 | 0.987 | 0.996 | 0.848 | 0.812 | 0.744 | 0.558 | 0.900 | 0.19 |
| 46 | 264 | 125 | 102 | 84 | 63 | 0.840 | 0.882 | 0.940 | 1.000 | 0.398 | 0.341 | 0.299 | 0.239 | 0.588 | 0.07 |
| 47 | 275 | 261 | 245 | 231 | 198 | 0.973 | 0.992 | 0.996 | 1.000 | 0.924 | 0.884 | 0.836 | 0.720 | 0.955 | 0.30 |
| 48 | 634 | 629 | 587 | 548 | 459 | 0.933 | 0.969 | 0.984 | 0.993 | 0.926 | 0.897 | 0.850 | 0.719 | 0.936 | 0.40 |
| 49 | 277 | 181 | 140 | 103 | 54 | 0.983 | 0.979 | 0.971 | 0.981 | 0.643 | 0.495 | 0.361 | 0.191 | 0.854 | 0.05 |
| 50 | 135 | 66 | 35 | 22 | 12 | 0.697 | 0.914 | 1.000 | 1.000 | 0.341 | 0.237 | 0.163 | 0.089 | 0.484 | 0.14 |
| 51 | 19 | 1 | 0 | 0 | 0 | 1.000 | NA | NA | NA | 0.053 | 0.000 | 0.000 | 0.000 | 0.261 | 0.12 |
| 52 | 112 | 50 | 26 | 17 | 4 | 0.840 | 0.962 | 1.000 | 1.000 | 0.375 | 0.223 | 0.152 | 0.036 | 0.598 | 0.14 |
| 53 | 268 | 110 | 69 | 44 | 21 | 0.827 | 0.957 | 0.977 | 1.000 | 0.340 | 0.246 | 0.160 | 0.078 | 0.543 | 0.10 |
| 54 | 37 | 1 | 0 | 0 | 0 | 1.000 | NA | NA | NA | 0.027 | 0.000 | 0.000 | 0.000 | 0.167 | 0.03 |
| 55 | 32 | 1 | 0 | 0 | 0 | 0.000 | NA | NA | NA | 0.000 | 0.000 | 0.000 | 0.000 | 0.212 | 0.03 |
| 56 | 700 | 575 | 477 | 397 | 273 | 0.816 | 0.881 | 0.927 | 0.963 | 0.670 | 0.600 | 0.526 | 0.376 | 0.741 | 0.28 |
| 57 | 20 | 1 | 0 | 0 | 0 | 1.000 | NA | NA | NA | 0.050 | 0.000 | 0.000 | 0.000 | 0.200 | 0.05 |
| 58 | 1413 | 1456 | 1200 | 857 | 422 | 0.788 | 0.848 | 0.939 | 0.995 | 0.812 | 0.720 | 0.570 | 0.297 | 0.802 | 0.28 |
| 59 | 40 | 14 | 9 | 4 | 1 | 0.429 | 0.444 | 0.500 | 1.000 | 0.150 | 0.100 | 0.050 | 0.025 | 0.429 | 0.06 |
| 60 | 283 | 135 | 87 | 63 | 34 | 0.741 | 0.851 | 0.873 | 0.824 | 0.353 | 0.261 | 0.194 | 0.099 | 0.492 | 0.18 |
| 61 | 944 | 943 | 802 | 641 | 388 | 0.832 | 0.899 | 0.944 | 0.985 | 0.832 | 0.764 | 0.641 | 0.405 | 0.840 | 0.33 |
| 62 | 330 | 227 | 186 | 163 | 131 | 0.877 | 0.930 | 0.957 | 0.977 | 0.603 | 0.524 | 0.473 | 0.388 | 0.722 | 0.20 |
| 63 | 421 | 339 | 287 | 239 | 184 | 0.873 | 0.923 | 0.958 | 0.995 | 0.703 | 0.629 | 0.544 | 0.435 | 0.784 | 0.24 |
| 64 | 778 | 816 | 747 | 696 | 632 | 0.876 | 0.924 | 0.953 | 0.978 | 0.919 | 0.887 | 0.852 | 0.794 | 0.907 | 0.46 |
| 65 | 286 | 233 | 217 | 209 | 172 | 0.927 | 0.959 | 0.962 | 0.988 | 0.755 | 0.727 | 0.703 | 0.594 | 0.838 | 0.41 |
| 66 | 261 | 166 | 136 | 104 | 79 | 0.843 | 0.934 | 0.962 | 0.975 | 0.536 | 0.487 | 0.383 | 0.295 | 0.673 | 0.12 |
| 67 | 135 | 60 | 29 | 8 | 0 | 0.800 | 0.931 | 1.000 | NA | 0.356 | 0.200 | 0.059 | 0.000 | 0.529 | 0.14 |

(*continued on next page*)

**Table 2** (*continued*)

| Class | In test set | Apparent detections | | | | Precision | | | | Recall | | | | Max F1 | At threshold |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| | | (0.25) | (0.50) | (0.75) | (0.95) | (0.25) | (0.50) | (0.75) | (0.95) | (0.25) | (0.50) | (0.75) | (0.95) | | |
| 68 | 279 | 286 | 271 | 256 | 244 | 0.916 | 0.952 | 0.973 | 0.984 | 0.939 | 0.925 | 0.892 | 0.860 | 0.943 | 0.54 |
| 69 | 85 | 61 | 45 | 40 | 29 | 0.770 | 0.933 | 0.950 | 1.000 | 0.553 | 0.494 | 0.447 | 0.341 | 0.667 | 0.15 |
| 70 | 27 | 4 | 0 | 0 | 0 | 0.250 | NA | NA | NA | 0.037 | 0.000 | 0.000 | 0.000 | 0.190 | 0.10 |
| 71 | 1719 | 1743 | 1447 | 1101 | 334 | 0.816 | 0.878 | 0.914 | 0.970 | 0.827 | 0.739 | 0.585 | 0.188 | 0.824 | 0.27 |
| 72 | 146 | 142 | 113 | 96 | 77 | 0.697 | 0.814 | 0.885 | 0.935 | 0.678 | 0.630 | 0.582 | 0.493 | 0.713 | 0.51 |
| 73 | 289 | 254 | 215 | 188 | 143 | 0.846 | 0.921 | 0.963 | 0.951 | 0.744 | 0.685 | 0.626 | 0.471 | 0.802 | 0.31 |
| 74 | 69 | 11 | 4 | 1 | 0 | 0.455 | 0.500 | 1.000 | NA | 0.072 | 0.029 | 0.014 | 0.000 | 0.196 | 0.12 |
| 75 | 58 | 30 | 24 | 17 | 6 | 0.700 | 0.750 | 0.765 | 0.833 | 0.362 | 0.310 | 0.224 | 0.086 | 0.511 | 0.11 |
| 76 | 718 | 705 | 687 | 678 | 658 | 0.966 | 0.975 | 0.981 | 0.988 | 0.948 | 0.933 | 0.926 | 0.905 | 0.958 | 0.22 |
| 77 | 156 | 121 | 44 | 11 | 0 | 0.455 | 0.614 | 0.818 | NA | 0.353 | 0.173 | 0.058 | 0.000 | 0.414 | 0.18 |
| 78 | 176 | 50 | 3 | 0 | 0 | 0.440 | 0.667 | NA | NA | 0.125 | 0.011 | 0.000 | 0.000 | 0.374 | 0.11 |
| 79 | 351 | 290 | 265 | 249 | 236 | 0.931 | 0.962 | 0.976 | 0.987 | 0.769 | 0.726 | 0.692 | 0.664 | 0.842 | 0.20 |
| 80 | 230 | 199 | 62 | 9 | 1 | 0.482 | 0.694 | 0.556 | 0.000 | 0.417 | 0.187 | 0.022 | 0.000 | 0.459 | 0.18 |
| 81 | 176 | 96 | 37 | 17 | 3 | 0.438 | 0.730 | 0.882 | 0.667 | 0.239 | 0.153 | 0.085 | 0.011 | 0.339 | 0.15 |
| 82 | 416 | 379 | 334 | 286 | 217 | 0.894 | 0.931 | 0.962 | 0.986 | 0.815 | 0.748 | 0.661 | 0.514 | 0.853 | 0.23 |
| 83 | 257 | 185 | 74 | 37 | 9 | 0.611 | 0.838 | 0.865 | 1.000 | 0.440 | 0.241 | 0.125 | 0.035 | 0.527 | 0.20 |
| 84 | 24 | 19 | 18 | 17 | 16 | 0.947 | 1.000 | 1.000 | 1.000 | 0.750 | 0.750 | 0.708 | 0.667 | 0.864 | 0.14 |
| 85 | 1680 | 1529 | 1316 | 1095 | 746 | 0.872 | 0.935 | 0.971 | 0.992 | 0.793 | 0.732 | 0.633 | 0.440 | 0.833 | 0.23 |
| 86 | 61 | 45 | 39 | 35 | 33 | 0.911 | 0.974 | 1.000 | 1.000 | 0.672 | 0.623 | 0.574 | 0.541 | 0.788 | 0.35 |
| 87 | 82 | 0 | 0 | 0 | 0 | NA | NA | NA | NA | 0.000 | 0.000 | 0.000 | 0.000 | 0.118 | 0.04 |
| 88 | 29 | 0 | 0 | 0 | 0 | NA | NA | NA | NA | 0.000 | 0.000 | 0.000 | 0.000 | 0.171 | 0.07 |
| 89 | 25 | 1 | 0 | 0 | 0 | 1.000 | NA | NA | NA | 0.040 | 0.000 | 0.000 | 0.000 | 0.427 | 0.03 |
| 90 | 583 | 510 | 389 | 304 | 179 | 0.763 | 0.869 | 0.921 | 0.961 | 0.667 | 0.580 | 0.480 | 0.295 | 0.717 | 0.35 |
| 91 | 488 | 212 | 97 | 60 | 26 | 0.675 | 0.784 | 0.917 | 0.962 | 0.293 | 0.156 | 0.113 | 0.051 | 0.461 | 0.11 |
| 92 | 57 | 23 | 12 | 6 | 2 | 0.826 | 0.917 | 0.833 | 1.000 | 0.333 | 0.193 | 0.088 | 0.035 | 0.583 | 0.06 |
| 93 | 31 | 1 | 0 | 0 | 0 | 1.000 | NA | NA | NA | 0.032 | 0.000 | 0.000 | 0.000 | 0.200 | 0.03 |
| 94 | 26 | 17 | 17 | 15 | 9 | 1.000 | 1.000 | 1.000 | 1.000 | 0.654 | 0.654 | 0.577 | 0.346 | 0.852 | 0.03 |
| 95 | 183 | 125 | 77 | 39 | 16 | 0.680 | 0.831 | 0.949 | 1.000 | 0.464 | 0.350 | 0.202 | 0.087 | 0.599 | 0.13 |
| 96 | 273 | 263 | 235 | 221 | 200 | 0.924 | 0.974 | 0.986 | 0.990 | 0.890 | 0.839 | 0.799 | 0.725 | 0.914 | 0.35 |
| 97 | 284 | 320 | 271 | 241 | 193 | 0.831 | 0.915 | 0.971 | 0.990 | 0.937 | 0.873 | 0.824 | 0.673 | 0.900 | 0.71 |
| 98 | 18 | 12 | 11 | 10 | 7 | 0.917 | 1.000 | 1.000 | 1.000 | 0.611 | 0.611 | 0.556 | 0.389 | 0.824 | 0.10 |
| 99 | 64 | 14 | 3 | 0 | 0 | 0.643 | 1.000 | NA | NA | 0.141 | 0.047 | 0.000 | 0.000 | 0.412 | 0.06 |
| 100 | 25 | 4 | 0 | 0 | 0 | 0.500 | NA | NA | NA | 0.080 | 0.000 | 0.000 | 0.000 | 0.418 | 0.03 |
| 101 | 27 | 25 | 23 | 19 | 18 | 0.880 | 0.913 | 1.000 | 1.000 | 0.815 | 0.778 | 0.704 | 0.667 | 0.857 | 0.52 |
| 102 | 31 | 29 | 24 | 22 | 16 | 0.828 | 0.875 | 0.909 | 1.000 | 0.774 | 0.677 | 0.645 | 0.516 | 0.800 | 0.24 |
| 103 | 24 | 29 | 14 | 2 | 0 | 0.310 | 0.429 | 1.000 | NA | 0.375 | 0.250 | 0.083 | 0.000 | 0.375 | 0.32 |
| 104 | 67 | 46 | 27 | 10 | 0 | 0.804 | 1.000 | 1.000 | NA | 0.552 | 0.403 | 0.149 | 0.000 | 0.673 | 0.29 |
| 105 | 70 | 55 | 26 | 15 | 6 | 0.527 | 0.808 | 0.867 | 1.000 | 0.414 | 0.300 | 0.186 | 0.086 | 0.485 | 0.21 |
| 106 | 92 | 59 | 50 | 45 | 31 | 1.000 | 1.000 | 1.000 | 1.000 | 0.641 | 0.543 | 0.489 | 0.337 | 0.851 | 0.10 |
| 107 | 28 | 5 | 3 | 1 | 0 | 1.000 | 1.000 | 1.000 | NA | 0.179 | 0.107 | 0.036 | 0.000 | 0.452 | 0.04 |
| 108 | 1075 | 1046 | 1001 | 965 | 870 | 0.925 | 0.949 | 0.965 | 0.985 | 0.900 | 0.884 | 0.866 | 0.797 | 0.917 | 0.57 |
| 109 | 275 | 225 | 208 | 196 | 175 | 0.938 | 0.971 | 0.985 | 1.000 | 0.767 | 0.735 | 0.702 | 0.636 | 0.845 | 0.32 |
| 110 | 20 | 12 | 9 | 7 | 1 | 1.000 | 1.000 | 1.000 | 1.000 | 0.600 | 0.450 | 0.350 | 0.050 | 0.750 | 0.16 |
| 111 | 767 | 811 | 743 | 694 | 623 | 0.837 | 0.896 | 0.928 | 0.966 | 0.885 | 0.868 | 0.840 | 0.785 | 0.885 | 0.67 |
| 112 | 726 | 706 | 646 | 600 | 491 | 0.887 | 0.924 | 0.950 | 0.986 | 0.862 | 0.822 | 0.785 | 0.667 | 0.879 | 0.39 |
| 113 | 37 | 23 | 21 | 18 | 15 | 0.913 | 0.952 | 1.000 | 1.000 | 0.568 | 0.541 | 0.486 | 0.405 | 0.727 | 0.07 |
| 114 | 25 | 17 | 12 | 8 | 3 | 0.824 | 0.917 | 1.000 | 1.000 | 0.560 | 0.440 | 0.320 | 0.120 | 0.694 | 0.05 |
| 115 | 453 | 254 | 115 | 52 | 11 | 0.626 | 0.843 | 0.923 | 0.909 | 0.351 | 0.214 | 0.106 | 0.022 | 0.483 | 0.12 |
| 116 | 242 | 163 | 122 | 99 | 56 | 0.847 | 0.926 | 0.949 | 0.964 | 0.570 | 0.467 | 0.388 | 0.223 | 0.695 | 0.12 |
| 117 | 57 | 38 | 35 | 30 | 13 | 1.000 | 1.000 | 1.000 | 1.000 | 0.667 | 0.614 | 0.526 | 0.228 | 0.899 | 0.03 |
| 118 | 318 | 291 | 268 | 241 | 176 | 0.897 | 0.940 | 0.975 | 0.994 | 0.821 | 0.792 | 0.739 | 0.550 | 0.867 | 0.59 |
| 119 | 253 | 235 | 118 | 54 | 3 | 0.566 | 0.805 | 0.963 | 1.000 | 0.526 | 0.375 | 0.206 | 0.012 | 0.556 | 0.29 |
| 120 | 32 | 2 | 0 | 0 | 0 | 1.000 | NA | NA | NA | 0.062 | 0.000 | 0.000 | 0.000 | 0.311 | 0.09 |
| 121 | 490 | 244 | 158 | 124 | 85 | 0.816 | 0.924 | 0.984 | 1.000 | 0.406 | 0.298 | 0.249 | 0.173 | 0.596 | 0.14 |
| 122 | 501 | 502 | 498 | 496 | 488 | 0.994 | 0.998 | 0.998 | 1.000 | 0.996 | 0.992 | 0.988 | 0.974 | 0.995 | 0.22 |
| 123 | 138 | 114 | 94 | 73 | 50 | 0.842 | 0.947 | 0.973 | 0.980 | 0.696 | 0.645 | 0.514 | 0.355 | 0.787 | 0.31 |
| 124 | 345 | 358 | 306 | 260 | 179 | 0.785 | 0.859 | 0.919 | 0.961 | 0.814 | 0.762 | 0.693 | 0.499 | 0.812 | 0.37 |
| 125 | 82 | 65 | 56 | 45 | 33 | 0.862 | 0.964 | 0.978 | 1.000 | 0.683 | 0.659 | 0.537 | 0.402 | 0.800 | 0.33 |
| 126 | 746 | 683 | 601 | 542 | 466 | 0.899 | 0.950 | 0.974 | 0.987 | 0.823 | 0.765 | 0.708 | 0.617 | 0.863 | 0.34 |
| 127 | 1968 | 1655 | 1215 | 966 | 708 | 0.804 | 0.915 | 0.968 | 0.986 | 0.676 | 0.565 | 0.475 | 0.355 | 0.735 | 0.21 |
| 128 | 220 | 117 | 73 | 46 | 25 | 0.726 | 0.849 | 0.935 | 1.000 | 0.386 | 0.282 | 0.195 | 0.114 | 0.523 | 0.18 |
| 129 | 221 | 140 | 69 | 30 | 7 | 0.700 | 0.812 | 0.867 | 1.000 | 0.443 | 0.253 | 0.118 | 0.032 | 0.562 | 0.19 |
| 130 | 465 | 408 | 349 | 318 | 252 | 0.900 | 0.946 | 0.965 | 0.992 | 0.789 | 0.710 | 0.660 | 0.538 | 0.842 | 0.24 |
| 131 | 1078 | 1052 | 735 | 543 | 303 | 0.729 | 0.869 | 0.947 | 0.987 | 0.712 | 0.593 | 0.477 | 0.277 | 0.723 | 0.27 |
| 132 | 3700 | 4071 | 2464 | 1584 | 796 | 0.724 | 0.890 | 0.961 | 0.991 | 0.797 | 0.593 | 0.412 | 0.213 | 0.763 | 0.30 |
| 133 | 73 | 70 | 55 | 47 | 21 | 0.957 | 0.982 | 1.000 | 1.000 | 0.918 | 0.740 | 0.644 | 0.288 | 0.939 | 0.11 |
| 134 | 816 | 856 | 720 | 571 | 321 | 0.761 | 0.838 | 0.891 | 0.960 | 0.798 | 0.739 | 0.624 | 0.377 | 0.792 | 0.40 |
| 135 | 2488 | 2526 | 2488 | 2450 | 2368 | 0.973 | 0.981 | 0.987 | 0.995 | 0.988 | 0.981 | 0.971 | 0.947 | 0.982 | 0.43 |

(Fig. 3). The shape of the relationship between threshold and apparent detections reflects the distribution of scores in the underlying dataset; a slower decline in the number of apparent detections suggests that scores are concentrated at or near 1, while a sharper decline suggests a more uniform distribution of scores, although the former case does not necessarily correspond to higher overall precision. These threshold-response profiles will guide future deployment strategies and post-processing filters for large-scale analyses as well as training priorities

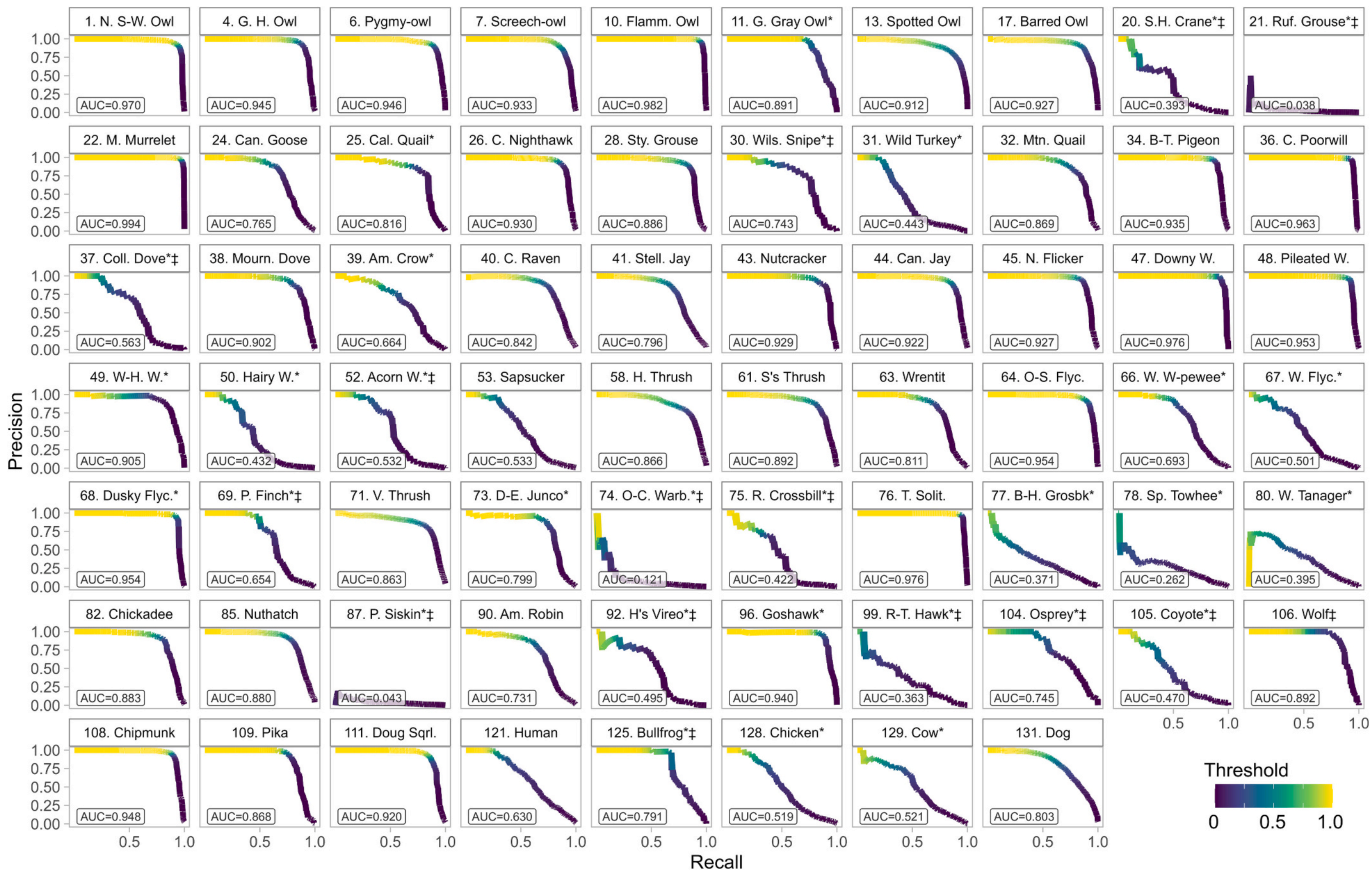


**Fig. 2.** Precision-recall curves for the primary target class for 68 vertebrate species detected by PNW-Cnet v5 based on a test set of 41,202 images which were held out from the full training dataset. Precision-recall curves plot precision (proportion of true positives among apparent detections) against recall (proportion of positive examples in the test set successfully detected) over a range of detection thresholds. AUC = area under the precision-recall curve, also known as mean average precision, used as a general measure of classifier performance. Precision-recall curves were not generated for classes with <40 positive examples in the test set. Classes are numbered as in Table 1, with abbreviated class descriptions for readability. An asterisk (*) indicates classes that were newly added in PNW-Cnet v5. A double dagger (‡) indicates classes that had <2000 images included in the training dataset for PNW-Cnet v5.

for future iterations of PNW-Cnet.

## 4. Discussion

The iterative development and application of PNW-Cnet represents a significant advancement in automated acoustic classification for ecological monitoring. By expanding from the six target classes of Ruff et al. (2020) to 135 target classes in PNW-Cnet v5, including 84 vertebrate species and a diverse set of anthropogenic and environmental sonotypes, the model supports broad-scale biodiversity assessment in complex forested landscapes of the western US. This expansion was achieved without compromising performance. The model maintained high accuracy even under conservative confidence score thresholds, illustrating the value of iterative model refinement and robust training strategies. These results align with previous findings that convolutional neural networks can generalize effectively across species and acoustic environments when trained on diverse and representative datasets (Kahl et al., 2021; Norman et al., 2022).

Beyond expanding the number of target classes, performance gains observed in legacy classes demonstrate the cumulative benefits of multi-year validation and retraining (Ruff et al., 2020; Ruff et al., 2021; Ruff et al., 2023). Although our test dataset differed from previous versions of PNW-Cnet, legacy classes showed higher and more stable precision and recall across thresholds, emphasizing the benefits of long-term data curation and iterative model updates. Due primarily to expanded training datasets, the successful incorporation of newly added classes, including rare and cryptic species, demonstrates the effectiveness of targeted data augmentation for addressing class imbalance and improving model generalization. While data augmentation is widely used in computer vision, it remains underutilized in bioacoustic applications. By simulating realistic acoustic conditions and increasing representation of rare classes, our augmentation framework enhanced model generalization and made it possible to include over two dozen species or environmental classes that would have otherwise lacked sufficient training data. These strategies are increasingly recognized as essential, due to the uneven distribution of sonotypes in real soundscapes and the extensive labor required for manual annotation (Stowell et al., 2019), which can otherwise limit the accessibility of deep learning models.

Unlike many published convolutional neural network-based classifiers, which often report only the list of included species, we have provided detailed performance metrics for each target class. This includes not only the number of training and test examples per class but also threshold-specific estimates of precision and recall. To our knowledge, this makes PNW-Cnet one of the only open-source convolutional neural networks supporting PAM programs that openly reports class-level performance metrics, filling a critical gap in transparency and reproducibility. Another important aspect of PNW-Cnet is its inclusion of non-biological classes alongside wildlife vocalizations. By explicitly modeling anthropogenic and environmental sounds, we enhanced model utility for varied use cases. These additional classes allow researchers and land managers to either filter out or retain non-biological sounds depending on their goals, facilitating more flexible and interpretable analyses. Incorporating “bioacoustic bycatch” has become a

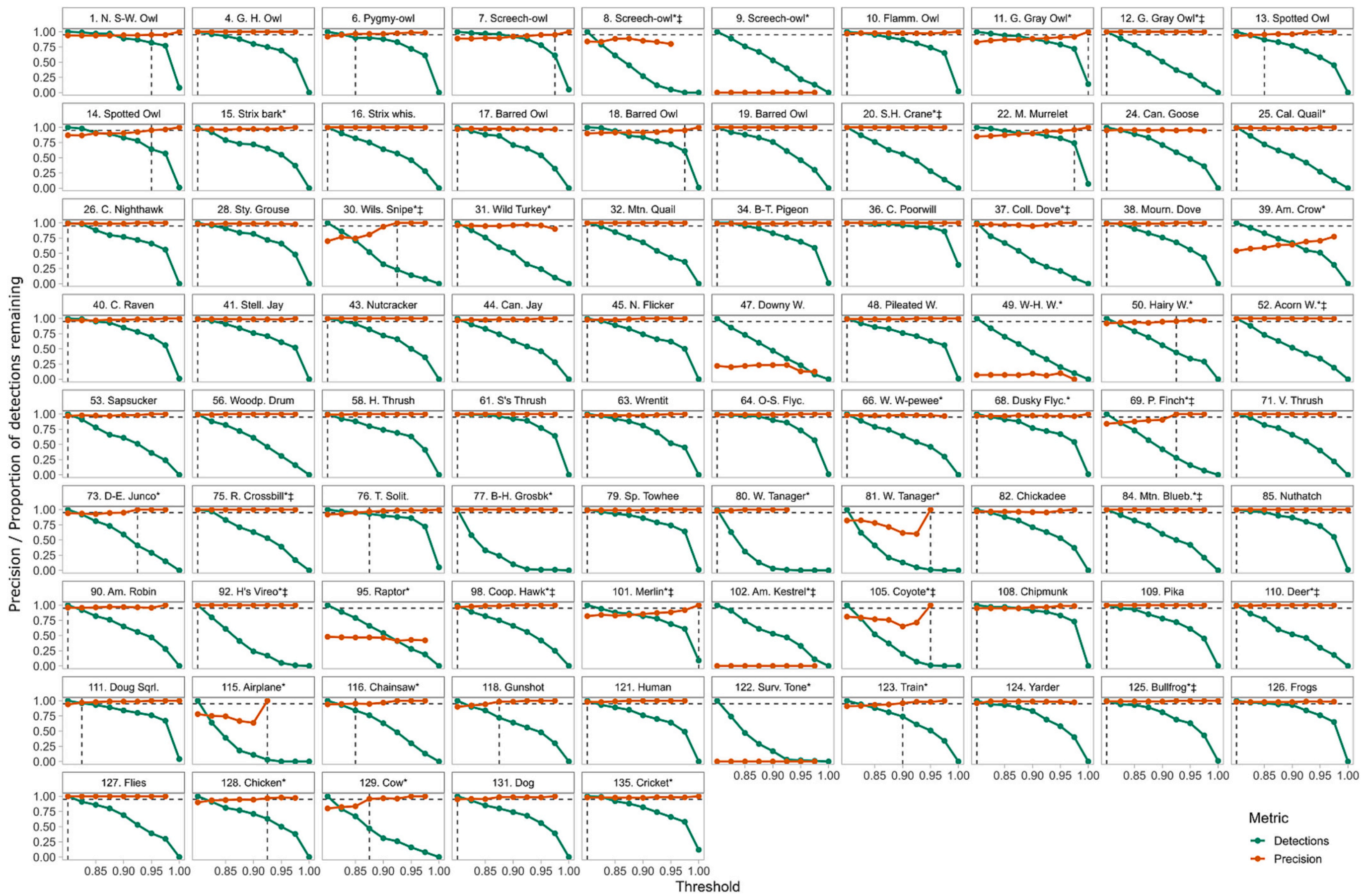


**Fig. 3.** Precision (orange) and proportion of detections remaining in the test dataset (green) at a range of thresholds for 85 PNW-Cnet v5 target classes. Precision was calculated based on a set of 8500 clips drawn from the 2023 data. We reviewed these clips fully, marking each clip as a true positive if it contained the target class and as a false positive if it did not. The test set initially included 100 clips for each class, selected randomly from the full set of clips from the 2023 data that met a detection threshold of 0.80. We then raised the threshold in increments of 0.025, and at each threshold, we calculated detections as the number of clips remaining as a proportion of the original 100 and calculated the proportion of remaining clips that were true positives. We then estimated the optimal threshold for each class as the lowest threshold (vertical dashed line) at which precision was ≥0.95 (horizontal dashed line). Some classes did not achieve precision ≥0.95 at any threshold. Classes are numbered as in Table 1, with abbreviated class descriptions for readability. An asterisk (*) indicates classes that were newly added in PNW-Cnet v5. A double dagger (‡) indicates classes that had <2000 images included in the training dataset for PNW-Cnet v5. (For interpretation of the references to colour in this figure legend, the reader is referred to the web version of this article.)

critical direction in PAM programs, particularly as land use change and human disturbance play increasingly prominent roles in shaping species distributions (Sugai et al., 2019).

Despite the strengths of PNW-Cnet v5, our estimates of precision based on the held-out test set may have been biased high. The training, validation, and held-out test set were random partitions of a single corpus of labeled images, which contained a high density of target classes compared to normal field recordings. As a result, the test set may not fully represent the acoustic heterogeneity present in field recordings, and precision metrics derived from it may overestimate the model's ability to distinguish target classes from background noise or unmodeled sonotypes. However, as noted above, PNW-Cnet v5 detects several common sources of environmental noise that were unmodeled in previous versions. Hence, the nature of the test set partly reflects the model's ability to capture an increased share of the acoustic complexity of the study area. Recall estimates from the held-out test set are likely more reliable, as the true number of positive examples for each class was known. This difference highlights the importance of interpreting test set metrics considering dataset structure and sampling strategy, and for robust validation protocols for producing data to use in ecological analyses.

The effect of class rarity was particularly pronounced in newly added species. For 26 classes, fewer than 50 examples were available in the test set, which may lead to precision and recall being overly sensitive to the characteristics of individual spectrograms. Low sample sizes for rare classes are a persistent challenge in multi-class bioacoustic models and may require additional augmentation, targeted data collection, few-shot learning, or transfer learning (Mcewen, 2024; Sethi et al., 2024; Sun et al., 2022; Weldy et al., 2025). We attempted to address this issue by augmenting our available training data for several newly added classes. The share of augmented training data varied widely by class, from <1% to >95%. Our approach appears to have been largely successful; most classes with a large proportion of augmented training data achieved good performance on one or both test sets (e.g., Great grey owl *Strix nebulosa*, Cooper's hawk), but performance was poorer for other species (e.g. Sandhill crane *Antigone canadensis*, American crow *Corvus brachyrhyncos*). Our approach to augmentation essentially relied on placing each unique example into multiple different acoustic contexts but did not introduce new variation in the signal itself. The usefulness of this approach depends at least partly on the initial set of training data and whether they cover the range of variation in the target class.

To complement the held-out test set, we conducted a field-based evaluation using apparent detections from the 2023 NWFP PAM program. This second test set better represented field conditions, including spatial, temporal, and acoustic variability. Because the number of positive examples in the full dataset was unknown, we focused on estimating empirical precision through manual review. This evaluation confirmed high precision for many classes, offset by consistent declines in apparent detections as thresholds increased, providing an indirect measure of recall loss. Although we could not directly estimate recall from this dataset, the trend illustrates a well-known trade-off in classifier behavior: higher thresholds reduce false positives but increase false negatives, while lower thresholds do the opposite. Threshold selection is therefore a critical decision point for users of automated classifiers like PNW-Cnet. Depending on study objectives, users may prioritize either recall or precision. In the case of rare or softly vocalizing species, or those in which one sex is especially vulnerable to omission at high thresholds, permissive thresholds (higher detection probability but increased review effort) or alternative screening strategies may be necessary (Appel et al., 2023). However, for occupancy models that assume no false positive observations to map species distributions, higher thresholds reduce potential false positive contaminations while streamlining data validation and allowing false negatives to be dealt with by directly estimating detection probability (Mackenzie et al., 2018). Recent work has begun to explore using score distributions or full model output for false-positive abundance models (Doser et al., 2022; Navine et al., 2024; Rhinehart et al., 2022; Udell et al., 2024), offering a potential path forward for abundance estimation not requiring binary detection decisions.

We filtered the field-based evaluation set based on estimated geographic range for each species to exclude apparent detections where occurrence is unlikely, which may have eliminated some sources of false positives, possibly resulting in higher precision than might be expected from a completely random sample. We consider this appropriate when assessing model performance. PNW-Cnet and similar models are deterministic and make predictions based entirely on the characteristics of spectrograms, which can lead to incorrect inferences if model output is accepted uncritically. We encourage practitioners to view model output skeptically and to incorporate other available information (e.g., knowledge of species geographic ranges) when interpreting these outputs, and especially when inferences based on PAM are used to guide management decisions.

In addition to model performance, operational deployment is a central consideration for convolutional neural networks developed to support PAM programs. PNW-Cnet v5 is integrated into both a graphical interface (Shiny_PNW-Cnet) and a command-line Python package (pycnet-audio), which allows users to apply the model flexibly and efficiently. These tools support the processing of large acoustic datasets on consumer-grade hardware and feature customizable class selection and threshold filtering so that users can tailor output to their specific research or management objectives. Similar efforts (e.g., Appel et al., 2025; Lapp et al., 2023) underscore the value of user-focused tools in operationalizing models to support scalable ecological inference.

Looking forward, further improvements to PNW-Cnet may include distinguishing sex of priority management species, such as spotted owls, which could enhance population monitoring by establishing pair status (Appel et al., 2023). Improvements may be achieved through transfer learning, where models trained on one region or taxonomic group are adapted to another; through cross-taxa or cross-ecosystem training; and through the integration of generative methods to produce high-quality synthetic training data for rare and cryptic species (Rafiq et al., 2025). Other promising directions include semi-supervised learning to leverage unannotated data, active learning to prioritize informative samples for labeling, and edge computing solutions that enable real-time classification (Mcloughlin et al., 2019). The flexible structure of PNW-Cnet makes it well suited to these future developments, and the model architecture and codebase are designed to support extension and retraining as monitoring needs evolve.

PNW-Cnet v5 illustrates how deep learning models can evolve in tandem with the ecological and management needs of a long-term monitoring program. Originating as a tool for detecting a small number of focal species, the model has expanded in both scope and capability through successive retraining on expert-labeled data and responsive class development. It now supports detection of over 80 wildlife species as well as diverse environmental and anthropogenic sound types, providing a foundation for multispecies inference and soundscape-level ecological assessment. Through deliberate model expansion, training data augmentation, rigorous evaluation, and integration with user-friendly software tools, PNW-Cnet serves as a central component of a reproducible and scalable framework for extracting ecological insight. As PAM programs continue to expand, tools like PNW-Cnet will be essential for enabling efficient, adaptable, and biologically informed monitoring of forest ecosystems and the species that depend on them. Beyond improved model performance, this progression demonstrates a shift in the role of PAM from narrowly focused species surveys to a scalable platform for ecosystem observation. By combining high-resolution acoustic data with adaptable, open-source modeling tools, we can support a new generation of biodiversity monitoring that is rigorous, repeatable, and regionally grounded (Tosa et al., 2021). As models like this continue to evolve, their value will depend not only on precision and recall but also on their ability to serve diverse stakeholders, respond to emerging conservation priorities, and contribute meaningfully to understanding and sustaining ecosystems.

## CRediT authorship contribution statement

**Damon B. Lesmeister:** Validation, Supervision, Resources, Project administration, Methodology, Investigation, Funding acquisition, Data curation, Conceptualization, Writing – review & editing, Writing – original draft. **Zachary J. Ruff:** Visualization, Validation, Software, Methodology, Investigation, Formal analysis, Data curation, Conceptualization, Writing – review & editing, Writing – original draft. **Adam Duarte:** Resources, Investigation, Data curation, Writing – review & editing. **Anna B. Kohlberg:** Validation, Methodology, Data curation, Writing – review & editing. **Taal Levi:** Supervision, Resources, Project administration, Funding acquisition, Conceptualization, Writing – review & editing. **Christopher M. Sullivan:** Supervision, Resources, Project administration, Funding acquisition, Conceptualization, Writing – review & editing.

## Funding sources

This work was supported by the U.S.D.I. Bureau of Land Management and U.S.D.A. Forest Service, with additional infrastructure and computational support provided by the Center for Quantitative Life Sciences at Oregon State University.

## Declaration of competing interest

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

## Acknowledgements

We thank the many field technicians, data reviewers, and collaborators who contributed to the U.S.D.A. Forest Service, Pacific Northwest Research Station's passive acoustic monitoring program from 2018 to 2023. Development and testing of PNW-Cnet v5 would not have been possible without the collective efforts of this team in funding acquisition, program management, deploying autonomous recorders, managing data infrastructure, and curating high-quality training data. This work was supported by the U.S.D.I. Bureau of Land Management and U.S.D.A. Forest Service, with additional infrastructure and computational support provided by the Center for Quantitative Life Sciences at Oregon State University. We also thank collaborators who provided feedback on early versions of this manuscript, the model, workflow, and the design and implementation of pycnet-audio and Shiny_PNW-Cnet. The findings

and conclusions in this article are those of the authors and do not necessarily represent the views of the U.S.D.A. Forest Service or other supporting institutions. The use of trade or firm names is for reader information only and does not imply endorsement by the U.S. Government.

## Appendix A. Supplementary data

Supplementary data to this article can be found online at https://doi.org/10.1016/j.ecoinf.2026.103657.

## Data availability

Our trained convolutional neural network, test data, and code sufficient to repeat the presented analyses are archived on Zenodo at DOI: https://doi.org/10.5281/zenodo.17955691.

## References


Abadi, M., Agarwal, A., Barham, P., Brevdo, E., Chen, Z., Citro, C., Corrado, G.S., Davis, A., Dean, J., Devin, M., Ghemawat, S., Goodfellow, I., Harp, A., Irving, G., Isard, M., Jozefowicz, R., 2015. *TensorFlow: Large-Scale Machine Learning on Heterogeneous Systems* [Online]. https://www.tensorflow.org/. Available: https://www.tensorflow.org/ [accessed].

Anderson, C.B., 2018. Biodiversity monitoring, earth observations and the ecology of scale. Ecol. Lett. 21, 1572–1585.

Appel, C.L., Lesmeister, D.B., Duarte, A., Davis, R.J., Weldy, M.J., Levi, T., 2023. Using passive acoustic monitoring to estimate northern spotted owl landscape use and pair occupancy. Ecosphere 14, e4421.

Appel, C.L., Subramanian, A., Koning, J.S., Ngosi, M., Sullivan, C.M., Levi, T., Lesmeister, D.B., 2025. Developing custom computer vision models with Njobvu-AI: A collaborative, user-friendly platform for ecological research. Ecol. Appl. 35, e70096.

Blumstein, D.T., Mennill, D.J., Clemins, P., Girod, L., Yao, K., Patricelli, G., Deppe, J.L., Krakauer, A.H., Clark, C., Cortopassi, K.A., Hanser, S.F., Mccowan, B., Ali, A.M., Kirschel, A.N.G., 2011. Acoustic monitoring in terrestrial environments using microphone arrays: applications, technological considerations and prospectus. J. Appl. Ecol. 48, 758–767.

Bota, G., Manzano-Rubio, R., Fanlo, H., Franch, N., Brotons, L., Villero, D., Devisscher, S., Pavesi, A., Cavaletti, E., Pérez-Granados, C., 2024. Passive acoustic monitoring and automated detection of the American bullfrog. Biol. Invasions 26, 1269–1279.

Burivalova, Z., Game, E.T., Butler, R.A., 2019. The sound of a tropical forest: recording of forest soundscapes can help monitor animal biodiversity for conservation. Science 363, 28–29.

Bush, A., Sollmann, R., Wilting, A., Bohmann, K., Cole, B., Balzter, H., Martius, C., Zlinszky, A., Calvignac-Spencer, S., Cobbold, C.A., Dawson, T.P., Emerson, B.C., Ferrier, S., Gilbert, M.T.P., Herold, M., Jones, L., Leendertz, F.H., Matthews, L., Millington, J.D.A., Olson, J.R., Ovaskainen, O., Raffaelli, D., Reeve, R., Rodel, M.O., Rodgers, T.W., Snape, S., Visseren-Hamakers, I., Vogler, A.P., White, P.C.L., Wooster, M.J., Yu, D.W., 2017. Connecting earth observation to high-throughput biodiversity data. Nat. Ecol. Evol. 1, 176.

Campos-Cerqueira, M., Aide, T.M., 2016. Improving distribution data of threatened species by combining acoustic monitoring and occupancy modelling. Methods Ecol. Evol. 7, 1340–1348.

Darras, K., Batáry, P., Furnas, B., Celis-Murillo, A., Van Wilgenburg, S.L., Mulyani, Y.A., Tscharntke, T., 2018. Comparing the sampling performance of sound recorders versus point counts in bird surveys: a meta-analysis. J. Appl. Ecol. 55, 2575–2586.

Darras, K., Batary, P., Furnas, B.J., Grass, I., Mulyani, Y.A., Tscharntke, T., 2019. Autonomous sound recording outperforms human observation for sampling birds: a systematic map and user guide. Ecol. Appl. 29, e01954.

Darras, K.F.A., Rountree, R.A., Van Wilgenburg, S.L., Cord, A.F., Pitz, F., Chen, Y., Dong, L., Rocquencourt, A., Desjonquères, C., Diaz, P.M., Lin, T.H., Turco, T., Emmerson, L., Bradfer-Lawrence, T., Gasc, A., Marley, S., Salton, M., Schillé, L., Wensveen, P.J., Wu, S.H., Acero-Murcia, A.C., Acevedo-Charry, O., Adam, M., Aguzzi, J., Akoglu, I., Amorim, M.C.P., Anders, M., André, M., Antonelli, A., Do Nascimento, L.A., Appel, G., Archer, S., Astaras, C., Atemasov, A., Atkinson, J., Attia, J., Baltag, E., Barbaro, L., Basan, F., Batist, C., Baumgarten, J.E., Bayle Sempere, J.T., Bellisario, K., David, A.B., Berger-Tal, O., Bertucci, F., Betts, M.G., Bhalla, I.S., Bicudo, T., Bolgan, M., Bombaci, S., Bota, G., Boullhesen, M., Briers, R. A., Buchan, S., Budka, M., Burchard, K., Buscaino, G., Calvente, A., Campos-Cerqueira, M., Gonçalves, M.I.C., Ceraulo, M., Cerezo-Araujo, M., Cerwén, G., Chaskda, A.A., Chistopolova, M., Clark, C.W., Cox, K.D., Cretois, B., Czarnecki, C., Da Silva, L.P., Da Silva, W., De Clippele, L.H., De La Haye, D., De Oliveira Tissiani, A. S., De Zwaan, D., Degano, M.E., Deichmann, J., Del Rio, J., Devenish, C., Díaz-Delgado, R., Diniz, P., Oliveira-Júnior, D.D., Dorigo, T., Dröge, S., Duarte, M., Duarte, A., Dunleavy, K., Dziak, R., Elise, S., Enari, H., Enari, H.S., Erbs, F., Eriksson, B.K., Ertör-Akyazi, P., Ferrari, N.C., Ferreira, L., Fleishman, A.B., Fonseca, P.J., Freitas, B., et al., 2025. Worldwide soundscapes: a synthesis of passive acoustic monitoring across realms. Glob. Ecol. Biogeogr. 34, e70021.

Doser, J.W., Leuenberger, W., Sillett, T.S., Hallworth, M.T., Zipkin, E.F., 2022. Integrated community occupancy models: A framework to assess occurrence and biodiversity dynamics using multiple data sources. Methods Ecol. Evol. 13, 919–932.

Duarte, A., Weldy, M.J., Lesmeister, D.B., Ruff, Z.J., Jenkins, J.M.A., Valente, J.J., Betts, M.G., 2024. Passive acoustic monitoring and convolutional neural networks facilitate high-resolution and broadscale monitoring of a threatened species. Ecol. Indic. 162, 112016.

Duchac, L.S., Lesmeister, D.B., Dugger, K.M., Ruff, Z.J., Davis, R.J., 2020. Passive acoustic monitoring effectively detects northern spotted owls and barred owls over a range of forest conditions. Condor 122, 1–22.

Duchac, L.S., Lesmeister, D.B., Dugger, K.M., Davis, R.J., 2021. Differential landscape use by forest owls two years after a mixed-severity wildfire. Ecosphere 12, e03770.

Farina, A., Krause, B., Mullet, T.C., 2024. An exploration of ecoacoustics and its applications in conservation ecology. BioSystems 245, 105296.

Gaylord, M.A., Duarte, A., Mccomb, B.C., Ratliff, J., 2023. Passive acoustic recorders increase White-headed woodpecker detectability in the Blue Mountains. J. Field Ornithol. 94, 1.

Gibb, R., Browning, E., Glover-Kapfer, P., Jones, K.E., Börger, L., 2019. Emerging opportunities and challenges for passive acoustics in ecological assessment and monitoring. Methods Ecol. Evol. 10, 169–185.

Kahl, S., Wood, C.M., Eibl, M., Klinck, H., 2021. BirdNET: A deep learning solution for avian diversity monitoring. Eco. Inform. 61, 101236.

Knight, E.C., Hannah, K.C., Foley, G., Scott, C., Brigham, R.M., Bayne, E., 2017. Recommendations for acoustic recognizer performance assessment with application to five common automated signal recognition programs. Avian Conserv. Ecol. 12, Article 14.

Lapp, S., Rhinehart, T., Freeland-Haynes, L., Khilnani, J., Syunkova, A., Kitzes, J., 2023. OpenSoundscape: an open-source bioacoustics analysis package for Python. Methods Ecol. Evol. 14, 2321–2328.

Lesmeister, D.B., Jenkins, J.M.A., 2022. Integrating new technologies to broaden the scope of northern spotted owl monitoring and linkage with USDA forest inventory data. Front. Forests Glob. Change 5, 966978.

Lesmeister, D.B., Davis, R.J., Singleton, P.H., Wiens, J.D., 2018. Northern spotted owl habitat and populations: Status and threats. In: Spies, T.A., Stine, P.A., Gravenmier, R., Long, J.W., Reilly, M.J. (Eds.), Synthesis of Science to Inform Land Management within the Northwest Forest Plan Area. General Technical Report PNW-GTR-966. USDA Forest Service, Pacific Northwest Research Station, Portland, OR.

Lesmeister, D.B., Appel, C.L., Davis, R.J., Yackulic, C.B., Ruff, Z.J., 2021. Simulating the Effort Necessary to Detect Changes in Northern Spotted Owl (*Strix occidentalis caurina*) Populations Using Passive Acoustic Monitoring. Research Paper PNW-RP-618. USDA Forest Service, Pacific Northwest Research Station, Portland, OR.

Lindenmayer, D.B., Likens, G.E., 2010. The science and application of ecological monitoring. Biol. Conserv. 143, 1317–1328.

Mackenzie, D.I., Nichols, J.D., Royle, J.A., Pollock, K.H., Bailey, L.L., Hines, J.E., 2018. Occupancy Estimation and Modeling: Inferring Patterns and Dynamics of Species Occurrence. Academic Press, Cambridge, MA.

Mcewen, B., 2024. Computational Bioacoustics for the Detection of Rare Acoustic Events. PhD, University of Canterbury.

Mcloughlin, M.P., Stewart, R., Mcelligott, A.G., 2019. Automated bioacoustics: methods in ecology and conservation and their potential for animal welfare monitoring. J. R. Soc. Interface 16, 20190225.

Morfi, V., Lachlan, R.F., Stowell, D., 2021. Deep perceptual embeddings for unlabelled animal sound events. J. Acoust. Soc. Am. 150, 2–11.

Navine, A.K., Denton, T., Weldy, M.J., Hart, P.J., 2024. All thresholds barred: direct estimation of call density in bioacoustic data. Front. Bird Sci. 3, 1380636.

Norman, D.L., Bischoff, P.H., Wearn, O.R., Ewers, R.M., Rowcliffe, J.M., Evans, B., Sethi, S., Chapman, P.M., Freeman, R., 2022. Can CNN-based species classification generalise across variation in habitat within a camera trap survey? Methods Ecol. Evol. 14, 242–251.

Pereira, H.M., Ferrier, S., Walters, M., Geller, G.N., Jongman, R.H.G., Scholes, R.J., Bruford, M.W., Brummitt, N., Butchart, S.H.M., Cardoso, A.C., Coops, N.C., Dulloo, E., Faith, D.P., Freyhof, J., Gregory, R.D., Heip, C., Höft, R., Hurtt, G., Jetz, W., Karp, D.S., Mcgeoch, M.A., Obura, D., Onoda, Y., Pettorelli, N., Reyers, B., Sayre, R., Scharlemann, J.P.W., Stuart, S.N., Turak, E., Walpole, M., Wegmann, M., 2013. Essential biodiversity variables. Science 339, 277–278.

Priyadarshani, N., Marsland, S., Castro, I., 2018. Automated birdsong recognition in complex acoustic environments: a review. J. Avian Biol. 49, e01447.

Rafiq, K., Beery, S., Palmer, M.S., Harchaoui, Z., Abrahms, B., 2025. Generative AI as a tool to accelerate the field of ecology. Nat. Ecol. Evol. 9, 378–385.

Rhinehart, T.A., Turek, D., Kitzes, J., 2022. A continuous-score occupancy model that incorporates uncertain machine learning output from autonomous biodiversity surveys. Methods Ecol. Evol. 13, 1778–1789.

Ross, S.R.P.J., O'connell, D.P., Deichmann, J.L., Desjonquères, C., Gasc, A., Phillips, J.N., Sethi, S.S., Wood, C.M., Burivalova, Z., 2023. Passive acoustic monitoring provides a fresh perspective on fundamental ecological questions. Funct. Ecol. 37, 959–975.

Ruff, Z.J., Lesmeister, D.B., Duchac, L.S., Padmaraju, B.K., Sullivan, C.M., 2020. Automated identification of avian vocalizations with deep convolutional neural networks. Rem. Sens. Ecol. Conserv. 6, 79–92.

Ruff, Z.J., Lesmeister, D.B., Appel, C.L., Sullivan, C.M., 2021. Workflow and convolutional neural network for automated identification of animal sounds. Ecol. Indic. 124, 107419.

Ruff, Z.J., Lesmeister, D.B., Jenkins, J.M.A., Sullivan, C.M., 2023. PNW-Cnet v4: automated species identification for passive acoustic monitoring. SoftwareX 23, 101473.

Rugg, N.M., Jenkins, J.M.A., Lesmeister, D.B., 2023. Western screech-owl occupancy in the face of an invasive predator. Glob. Ecol. Conserv. 48, e02753.

Sethi, S.S., Ewers, R.M., Jones, N.S., Signorelli, A., Picinali, L., Orme, C.D.L., Zamora-Gutierrez, V., 2020. SAFE acoustics: an open-source, real-time eco-acoustic monitoring network in the tropical rainforests of Borneo. Methods Ecol. Evol. 11, 1182–1185.
Sethi, S.S., Bick, A., Chen, M., Crouzeilles, R., Hillier, B.V., Lawson, J., Lee, C., Liu, S., Henrique De Freitas, C.P.C., Rosten, C.M., Somveille, M., Tuanmu, M., Banks-Leite, C., 2024. Large-scale avian vocalization detection delivers reliable global biodiversity insights. Proc. Natl. Acad. Sci. 121, e2315933121.
Spies, T.A., Long, J.W., Charnley, S., Hessburg, P.F., Marcot, B.G., Reeves, G.H., Lesmeister, D.B., Reilly, M.J., Cerveny, L.K., Stine, P.A., Raphael, M.G., 2019. Twenty-five years of the northwest Forest plan: what have we learned? Front. Ecol. Environ. 17, 511–520.
Stowell, D., 2022. Computational bioacoustics with deep learning: a review and roadmap. PeerJ 10, e13152.
Stowell, D., Wood, M.D., Pamuła, H., Stylianou, Y., Glotin, H., Orme, D., 2019. Automatic acoustic detection of birds through deep learning: the first bird audio detection challenge. Methods Ecol. Evol. 10, 368–380.
Sugai, L.S.M., Silva, T.S.F., Ribeiro Jr., J.W.R., Llusia, D., 2019. Terrestrial passive acoustic monitoring: review and perspectives. BioScience 69, 15–25.
Sun, Y., Midori Maeda, T., Solís-Lemus, C., Pimentel-Alarcón, D., Buřivalová, Z., 2022. Classification of animal sounds in a hyperdiverse rainforest using convolutional neural networks with data augmentation. Ecol. Indic. 145, 109621.
Tosa, M.I., Dziedzic, E.H., Appel, C.L., Urbina, J., Massey, A., Ruprecht, J., Eriksson, C.E., Dolliver, J.E., Lesmeister, D.B., Betts, M.G., Peres, C.A., Levi, T., 2021. The rapid rise of next-generation natural history. Front. Ecol. Evol. 9, 698131.
Udell, B.J., Straw, B.R., Loeb, S.C., Irvine, K.M., Thogmartin, W.E., Lausen, C.L., Reichard, J.D., Coleman, J.T.H., Cryan, P.M., Frick, W.F., Reichert, B.E., 2024. Using mobile acoustic monitoring and false-positive N-mixture models to estimate bat abundance and population trends. Ecol. Monogr. 94, e1617.
Van Parijs, S.M., Clark, C.W., Sousa-Lima, R.S., Parks, S.E., Rankin, S., Risch, D., Van Opzeeland, I.C., 2009. Management and research applications of real-time and archival passive acoustic sensors over varying temporal and spatial scales. Mar. Ecol. Prog. Ser. 395, 21–36.
Watson, M., Chollet, F.O., Sreepathihalli, D., Saadat, S., Sampath, R., Rasskin, G., Zhu, S., Singh, V., Wood, L., Tan, Z., Stenbit, I., Qian, C., Bischof, J., 2024. *KerasHub* [Online] [accessed].
Weldy, M.J., Lesmeister, D.B., Denton, T., Duarte, A., Vernasco, B.J., Gasc, A., Rowe, J. C., Adams, M.J., Betts, M.G., 2025. Simulated soundscapes and transfer learning boost the performance of acoustic classifiers under data scarcity. Methods Ecol. Evol. 17, 322–338.
Yoccoz, N.G., Nichols, J.D., Boulinier, T., 2001. Monitoring of biological diversity in space and time. Trends Ecol. Evol. 16, 446–453.