

Validação preliminar da eficácia do modelo Birdnet com dados do Pantanal

Leandro Moreira de Carvalho¹, Bianca Mitie Nakazawa¹, Rakell Aparecida Barbosa Guimarães¹, Lucas Korzune², Thiago Meirelles Ventura¹, Allan Gonçalves de Oliveira¹

¹ Instituto de Computação – Universidade Federal de Mato Grosso (UFMT) – Cuiabá – MT – Brazil

² Instituto Nacional de Pesquisas do Pantanal (INPP)

{leandro.carvalho, bianca.nakazawa}@sou.ufmt.br,
rakell.rafaelly@gmail.com, {thiago, allan}@ic.ufmt.br,
lucaskorzune@gmail.com

Abstract. *This study validates BirdNET for the Pantanal biome by analyzing over 500 species via Xeno-Canto. Utilizing top-1 and top-5 accuracy metrics, the model demonstrated robust identification performance despite challenges posed by dialectal variations and taxonomic divergences. Although 40 species lack specific model classes, the results confirm BirdNET's viability for the biome, with fine-tuning recommended for regional optimization.*

Resumo. *Este trabalho valida o BirdNET para o Pantanal analisando mais de 500 espécies via Xeno-Canto. Utilizando métricas de top-1 e top-5 accuracy, observou-se robustez na identificação, apesar de desafios com variações dialetais e divergências taxonômicas. Embora 40 espécies careçam de classes específicas, os resultados confirmam a viabilidade do modelo para o bioma, recomendando-se fine-tuning para otimização regional.*

1. Introdução

O BirdNET é um classificador bioacústico originalmente desenvolvido para a identificação automática de espécies de aves, treinado inicialmente com mais de 150.000 amostras de áudio de 984 espécies presentes na América do Norte, majoritariamente provenientes das bases eBird e Xeno-Canto [Kahl et al. 2021, eBird 2021, Xeno-canto Foundation 2024]. Posteriormente, o modelo foi expandido para uma escala global, passando a contemplar 6.522 classes acústicas, incluindo sons não biológicos e vocalizações de outros grupos faunísticos, como anfíbios anuros.

Estudos indicam que o desempenho do BirdNET varia conforme o bioma, apresentando resultados inferiores em áreas úmidas altamente biodiversas quando comparadas a florestas tropicais e subtropicais [Funosas et al. 2026]. Esse comportamento é particularmente relevante no contexto do Pantanal, a maior planície sazonalmente alagável do planeta, bioma megadiverso e submetido a crescentes pressões antrópicas, o que demanda estratégias eficientes de monitoramento ambiental [Junk et al. 2006, Thielen et al. 2020].

Embora o BirdNET inclua espécies do Pantanal em seu conjunto de classes pré-treinadas, o modelo não foi treinado com gravações locais, o que pode afetar seu desempenho regional. Nesse contexto, este trabalho apresenta resultados iniciais de uma análise de classificação acústica de aves do Pantanal utilizando o modelo baseline BirdNET

2. Metodologia

O fluxo de trabalho geral foi estruturado em três etapas principais: (I) aquisição dos dados para análise por meio de uma API; (II) aplicação de filtros quanto à quantidade de áudios, visando garantir uma comparação justa entre as espécies; e (III) análise sistemática dos resultados utilizando as métricas de desempenho *top-1 accuracy* e *top-5 accuracy*

Como critério inicial, foram selecionados cinco áudios por espécie para compor o conjunto de análise. Nos casos em que esse número de registros não estava disponível na base de dados online utilizada para a obtenção dos áudios, adotou-se o pareamento da quantidade de amostras nas análises específicas, de modo a assegurar o equilíbrio estatístico entre as espécies e categorias comparadas.

2.1. Aquisição de dados e Lista de espécies

Os dados foram obtidos por meio da API do Xeno-Canto [Xeno-canto Foundation 2024], considerando apenas registros de alta qualidade (classes A e B). Como o levantamento interno conduzido por um membro do grupo Computational Bioacoustics Research Unit [CO.BRA 2026] adota o padrão taxonômico IOC [Gill et al. 2024], enquanto o BirdNET utiliza o padrão Clements [Clements et al. 2023], realizou-se um mapeamento prévio de sinônimos para mitigar inconsistências taxonômicas. Ao todo, foram consideradas 617 espécies de aves registradas no Pantanal.

A amostra final foi ajustada de acordo com a disponibilidade de registros no Xeno-Canto, uma vez que nem todas as espécies apresentavam o número ideal de cinco áudios. As gravações utilizadas caracterizam-se predominantemente por ambientes naturais, com presença significativa de ruídos de fundo oriundos de fontes orgânicas e não orgânicas.

2.2. Critérios de análise

A avaliação de desempenho foi organizada em duas abordagens: uma análise generalista, abrangendo todas as espécies pantaneiras passíveis de identificação pelo BirdNET, e análises específicas por subconjuntos. Além disso, para a métrica *top-5 accuracy*, foi definido um limiar mínimo de confiança (*threshold*) de 0,05, visando filtrar predições com baixo grau de confiabilidade estatística.

3. Resultados e discussões

3.1. Análise de desempenho para as espécies pantaneiras identificadas pelo modelo BirdNET

A primeira etapa do estudo consistiu na avaliação do desempenho geral do BirdNET a partir de três abordagens complementares: (I) análise de registros coletados no Pantanal e em suas zonas de transição, delimitados geograficamente por meio do recurso de bounding box da API do Xeno-Canto [Xeno-canto Foundation 2024]; (II) análise de registros globais das espécies-alvo, sem restrição geográfica, visando ampliar o volume amostral;

e (III) seleção manual de 20 espécies, com cinco áudios representativos por espécie, para uma avaliação detalhada da precisão.

Os resultados dessas análises são apresentados na Tabela 1, que sumariza o volume de dados, as métricas de acurácia *top-1* e *top-5*, bem como a confiança média associada às predições corretas e incorretas.

Tabela 1. Desempenho comparativo do modelo BirdNET.

Natureza do Teste	Espécies	Áudios	Top-1	Top-5	Acertos	Erros
Áudios no Pantanal	366	1284	79,60%	88,16%	0,9236	0,6554
Espécies do Pantanal	549	2691	67,34%	81,38%	0,980	0,648
Separação manual	20	100	79,00%	88,00%	0,9220	0,5312

Os resultados apresentados na Tabela 1, indicam acurácia consistente entre as abordagens I e III (79%), enquanto a abordagem II apresentou redução para 67,34%, possivelmente em função da maior variabilidade acústica dos registros globais. Ainda assim, a confiança média nas predições corretas manteve-se elevada ($> 0,92$), atingindo 0,98 no cenário global, o que reforça a robustez das inferências corretas. Os erros com alta confiança ($> 0,70$) observados em registros individuais concentraram-se, em geral, em espécies taxonomicamente próximas e com vocalizações similares, como aquelas do gênero *Synallaxis*.

3.2. Análises específicas

Além da avaliação do desempenho geral do BirdNET em áudios de espécies do Pantanal, foram conduzidas análises específicas com foco comparativo. As análises realizadas foram: (I) comparação do desempenho do modelo entre áudios gravados na América do Norte e no Pantanal; (II) comparação entre áudios da América do Norte e Central versus áudios do Pantanal; (III) comparação entre áudios da Europa e do Pantanal; e (IV) análise das inferências do modelo em espécies do Pantanal não contempladas em suas classes, identificando os padrões de confusão taxonômica.

Os resultados das análises I–III estão sumarizados na Tabela 2, que apresenta as mesmas métricas descritas na Tabela 1. A análise IV é discutida em uma seção dedicada subsequente.

Tabela 2. Resultados das análises comparativas regionais do BirdNET.

Exp.	Região	Espécies	Áudios	Top-1	Top-5	Acertos	Erros
I	Pantanal	63	206	79,61%	90,78%	0,9135	0,6280
	Am. do Norte	63	206	78,64%	88,83%	0,9180	0,6102
II	Pantanal	197	626	81,63%	91,85%	0,9120	0,6454
	Am. Central/Norte	197	626	75,24%	86,90%	0,9157	0,6120
III	Pantanal	10	29	79,31%	89,66%	0,8891	0,7917
	Europa	10	29	72,41%	79,31%	0,8911	0,8357

Embora todas as análises tenham apresentado resultados satisfatórios, observou-se uma diferença sistemática de desempenho entre áudios gravados no Pantanal e aqueles

provenientes de outras regiões, com taxas de acerto no Pantanal ligeiramente superiores ou muito próximas. Esse comportamento sugere dois cenários explicativos principais.

(I) O modelo pode apresentar sensibilidade a variações dialetais regionais, conforme evidenciado por estudos clássicos que demonstram alterações na assinatura acústica de uma mesma espécie em função da localização geográfica [Marler and Tamura 1962], além de evidências de que o desempenho do BirdNET varia espacialmente [Kahl et al. 2021].

(II) Divergências taxonômicas entre as fontes de dados podem introduzir erros classificados como “falsos”. Enquanto o Xeno-Canto adota o padrão IOC, o BirdNET segue o padrão Clements/eBird, o que tende a gerar inconsistências, especialmente em regiões de elevada biodiversidade e frequentes revisões taxonômicas, como a América Central.

3.3. Ponderações em relação à análise IV

Na análise IV, avaliou-se o desempenho do modelo em 53 espécies inicialmente classificadas como biologicamente não reconhecíveis pelo BirdNET. Após uma análise detalhada das divergências entre os padrões taxonômicos e dos splits utilizados, verificou-se que 13 dessas espécies haviam sido erroneamente rotuladas.

Para as 40 espécies restantes, confirmou-se que o BirdNET não possui classes específicas para sua identificação. Ainda assim, observou-se um padrão consistente: mesmo na ausência da classe correta, o modelo frequentemente associou essas espécies a outras do mesmo gênero ou família. Esse comportamento indica que, apesar de não ter sido treinado com gravações específicas do Pantanal, o BirdNET é capaz de capturar características acústicas compartilhadas entre grupos taxonomicamente próximos, evidenciando uma robustez relevante na identificação em nível suprataxonômico.

4. Conclusão

O BirdNET demonstrou-se uma ferramenta eficaz para o monitoramento bioacústico no Pantanal, alcançando acurácia *top-1* entre 67% e 82%, com desempenho superior a 78% em dados locais. As principais limitações identificadas referem-se a divergências taxonômicas e à ausência de espécies regionais no conjunto de classes do modelo.

Como trabalhos futuros, recomenda-se a adoção sistemática de mapeamentos taxonômicos, bem como a aplicação de técnicas de *fine-tuning* ou *transfer learning* para aprimorar o desempenho em dados regionais. O treinamento de camadas específicas pode permitir a inclusão de espécies ausentes e a mitigação de variações dialetais regionais, aumentando a precisão do modelo. Adicionalmente, sugere-se a exploração de estratégias avançadas de pós-processamento e análise das inferências produzidas pelo BirdNET.

Agradecimentos

Agradecemos ao Computational Bioacoustics Research Unit (CO.BRA) da UFMT pelo apoio e orientação ao longo desta pesquisa

Referências

Clements, J. F., Schulenberg, T. S., Iliff, M. J., Fredericks, T. A., Gerbracht, J. A., Lepage, D., Billerman, S. M., Sullivan, B. L., and Wood, C. L. (2023). *The Clements Checklist*

of Birds of the World: v2023. Cornell Lab of Ornithology. Disponível em formato digital.

CO.BRA (2026). Computational bioacoustics research unit. Acessado em: 27 jan. 2026.

eBird (2021). ebird: An online database of bird distribution and abundance. Acessado em: 27 de jan. 2026.

Funosas, D., Sebastián-González, E., Morant, J., Marín Gómez, O. H., Mendoza, I., Mohedano-Muñoz, M. A., Santamaría, E., Bastianelli, G., Márquez-Rodríguez, A., Budka, M., Bota, G., Alonso-Moya, C. D., de la Peña-Rubio, J. M., de la Morena, E. L. G., Santa-Cruz, M., de la Nava, P., Fernández-Tizón, M., Sánchez-Mateos, H., Barrero, A., Traba, J., Osiejuk, T. S., Hart, P. J., Navine, A. K., Montoya Muñoz, A. F., de Araújo, C. B., Rosa, G. L., Torres, I. M., Catalano, A. L., Simões, C. R., Llusia, D., Morales, M. B., Acebes, P., Medina, J. A., Brown, N., Astaras, C., Karmiris, I., Navarrete, E., Cauchoix, M., Barbaro, L., Arend, D., Müeller, S., González-García, F., González-Romero, A., Mammides, C., Pontikis, M., Jacuzzi, G., Olden, J. D., Bombaci, S. P., Marcacci, G., Jacot, A., Zurano, J. P., Gangenova, E., Varela, D., Di Sallo, F., Zurita, G. A., Atemasov, A., Tremblay, J. A., Lamarre, V., Hutschenreiter, A., Monroy-Ojeda, A., Díaz-Vallejo, M., Chaparro-Herrera, S., Briers, R. A., Sousa-Lima, R., Pinheiro, T., Da Silva, W. C., Calvente, A., Paz, R. V., Salustio-Gomes, C., Oliveira-Júnior, D. D., Lima Santos, C. S., Pichorim, M., Molin, A. D., Antonelli, A., Gogoleva, S., Palko, I., Trong, H. V., Duarte, M. H., dos Santos Saturnino, N., Silva, S. R., Rainho, A., Lopes, P., Schuchmann, K.-L., Marques, M. I., de Oliverira Tissiani, A. S., Littlewood, N. A., Tuanmu, M.-N., Kepfer-Rojas, S., Aguilera, A. L., Brotons, L., Feldman, M. J., Imbeau, L., Panwar, P., Weed, A. S., Dehwal, A., Attisano, A., Theuerkauf, J., Goodale, E., Darras, K. F., and Pérez-Granados, C. (2026). A global assessment of birdnet performance: Differences among continents, biomes, and species. *Ecological Indicators*, 182:114550.

Gill, F., Donsker, D., and Rasmussen, P. (2024). IOC World Bird List (v14.1). Acessado em: 27 de jan. 2026.

Junk, W. J., da Cunha, C. N., Wantzen, K. M., Petermann, P., Strüssmann, C., Marques, M. I., and Adis, J. (2006). Biodiversity and its conservation in the pantanal of mato grosso, brazil. *Aquatic Sciences*, 68:278–309.

Kahl, S., Wood, C. M., Eibl, M., and Klinck, H. (2021). Birdnet: A deep learning solution for avian diversity monitoring. *Ecological Informatics*, 61:101236.

Marler, P. and Tamura, M. (1962). Song “dialects” in three populations of white-crowned sparrows. *The Condor: Ornithological Applications*, 64(5):368–377.

Thielen, D., Schuchmann, K.-L., Ramoni-Perazzi, P., Marquez, M., Rojas, W., Quintero, J. I., and Marques, M. I. (2020). Quo vadis pantanal? expected precipitation extremes and drought dynamics from changing sea surface temperature. *PLOS One*, 15(1):e0227437.

Xeno-canto Foundation (2024). Xeno-canto: Sharing bird sounds from around the world. Acessado em: 27 de jan. 2026.