

Assessing the Potential of Deep Learning in the Acoustic Identification of Threatened Birds in the Pantanal

Gabriel A. Correia¹, Thiago M. Ventura¹

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

`gabriel.acassio.correia@gmail.com, thiago@ic.ufmt.br`

Abstract. *This study investigates the feasibility of using machine learning techniques applied to bioacoustics for the automated identification of threatened bird species in the Pantanal biome. Bioacoustic data were collected from the Xeno-canto platform and used in conjunction with convolutional neural networks and Mel spectrograms. The results demonstrated satisfactory performance, with recall values varying among the selected species: *Anodorhynchus hyacinthinus* (0.94), *Calidris canutus* (0.93), *Calidris pusilla* (0.80), *Crax fasciolata* (0.70), and *Harpia harpyja* (0.89). It was concluded that the application of machine learning for detecting threatened bird species is technically feasible, depending on data quality and methodological adjustments.*

Resumo. *Este trabalho investiga a viabilidade do uso de técnicas de machine learning aplicadas à bioacústica para identificação automatizada de aves ameaçadas no bioma Pantanal. Dados bioacústicos foram coletados da plataforma Xeno-canto e utilizados juntamente com redes neurais convolucionais e mel spectrogram. Os resultados demonstraram desempenho satisfatório, com valores de recall que variaram entre as espécies selecionadas: *Anodorhynchus hyacinthinus* (0,94), *Calidris canutus* (0,93), *Calidris pusilla* (0,80), *Crax fasciolata* (0,70) e *Harpia harpyja* (0,89). Concluiu-se que a aplicação de machine learning para detecção de aves ameaçadas é tecnicamente viável, mesmo para espécies em extinção.*

1. Introdução

A biodiversidade constitui um dos pilares fundamentais para a saúde dos ecossistemas, desempenhando papel crucial na manutenção dos serviços ambientais que sustentam a vida no planeta. O Pantanal, reconhecido como uma das maiores áreas alagáveis do mundo, destaca-se como um verdadeiro santuário de diversidade biológica, abrigando uma variedade ímpar de espécies de flora e fauna [Junk et al. 2006]. No entanto, essa riqueza natural enfrenta ameaças crescentes provocadas por ações antropogênicas que comprometem a integridade dos habitats e afetam diretamente a resiliência das espécies.

Entre os grupos biológicos mais impactados, as aves ocupam lugar de destaque, sendo muitas delas classificadas em categorias de ameaça conforme os critérios da Lista Vermelha da IUCN [IUCN 2025] e do Livro Vermelho da Fauna Brasileira [Instituto Chico Mendes de Conservação da Biodiversidade 2018]. Segundo

[Instituto Chico Mendes de Conservação da Biodiversidade 2018], em 2018 havia aproximadamente 234 táxons de aves ameaçados de extinção no Brasil, dos quais 13 eram encontrados no Pantanal. As principais causas da redução populacional dessas espécies são a expansão urbana, a intensificação da agropecuária e a incidência de queimadas.

A conservação da biodiversidade exige estratégias de monitoramento eficazes. Nesse contexto, o monitoramento acústico passivo (*Passive Acoustic Monitoring*) revela-se uma alternativa promissora para o acompanhamento remoto da fauna, especialmente em áreas de difícil acesso. Quando associada a algoritmos de *Machine Learning* (ML), essa abordagem oferece meios eficientes para a identificação de espécies e a estimativa de suas populações.

A justificativa deste trabalho está ancorada na urgência de se desenvolver métodos eficazes voltados à conservação das aves ameaçadas do Pantanal. A adoção de ferramentas tecnológicas pode viabilizar monitoramentos mais precisos, menos invasivos e com maior cobertura espacial e temporal, preenchendo lacunas informacionais relevantes sobre as populações de aves em risco.

Diante desse cenário, este trabalho propõe explorar o uso de técnicas de ML, especificamente *deep learning*, aplicadas à bioacústica com o objetivo de identificar aves ameaçadas no Pantanal. O objetivo geral é verificar a viabilidade do emprego de técnicas de ML e de técnicas de tratamento e aumento de dados combinadas à bioacústica para a identificação de aves ameaçadas de extinção no Pantanal.

2. Fundamentação Teórica

2.1. Espécies Ameaçadas e Sua Relevância para a Conservação

Uma *espécie ameaçada* é aquela que enfrenta um risco significativo de extinção devido a fatores como destruição de habitat, caça, mudanças climáticas ou competição com espécies invasoras. A identificação de espécies ameaçadas é vital para a formulação de políticas e ações de conservação eficazes [Instituto Chico Mendes de Conservação da Biodiversidade 2018]. No Pantanal, as espécies são particularmente vulneráveis devido à degradação ambiental e às mudanças no uso do solo [Junk et al. 2006].

O monitoramento contínuo dessas espécies é crucial para prevenir a extinção e garantir a preservação do ecossistema pantaneiro. A bioacústica desempenha um papel importante ao permitir o rastreamento remoto de vocalizações, facilitando a coleta de dados sobre populações ameaçadas sem intervenções diretas no habitat natural [Sharma et al. 2022].

As listas vermelhas são ferramentas fundamentais no monitoramento da biodiversidade, incluindo a Lista Vermelha da IUCN e o Livro Vermelho da Fauna Brasileira. Estas listas categorizam as espécies com base no grau de risco de extinção: Pouco Preocupante (LC), Quase Ameaçada (NT), Vulnerável (VU), Em Perigo (EN), Criticamente em Perigo (CR), Extinta na Natureza (EW), Extinta (EX), Dados Insuficientes (DD) e Não Avaliada (NE). Essas categorias são determinadas com base em critérios como redução da população, perda de habitat e área de distribuição da espécie. As espécies em categorias de maior risco exigem monitoramento intensivo e esforços direcionados à conservação.

2.2. Bioacústica e Tratamento de Dados Acústicos

A bioacústica, o estudo dos sons produzidos por organismos vivos, constitui uma ferramenta poderosa para o monitoramento ambiental [Stowell 2022]. A bioacústica possibilita a detecção de espécies por meio de suas vocalizações, sem necessidade de observação visual. Microfones distribuídos por grandes áreas, realizando gravações contínuas, combinados com ML, permitem uma abordagem escalável e automática [de Oliveira et al. 2020].

Na análise de dados acústicos, os sinais de áudio são convertidos em representações visuais conhecidas como espectrogramas. Um espectrograma é uma imagem bidimensional onde o eixo horizontal representa o tempo, o eixo vertical a frequência, e a intensidade do som é codificada por uma escala de cores [Xie et al. 2019].

Uma variante comumente utilizada é o *mel spectrogram*, exemplificado pela Figura 1, que aplica uma transformação para representar as frequências segundo a escala de Mel, baseada na percepção humana de altura sonora. O resultado é uma matriz tempo-frequência com maior resolução em baixas frequências, favorecendo a detecção de características acústicas relevantes [Xie et al. 2019].

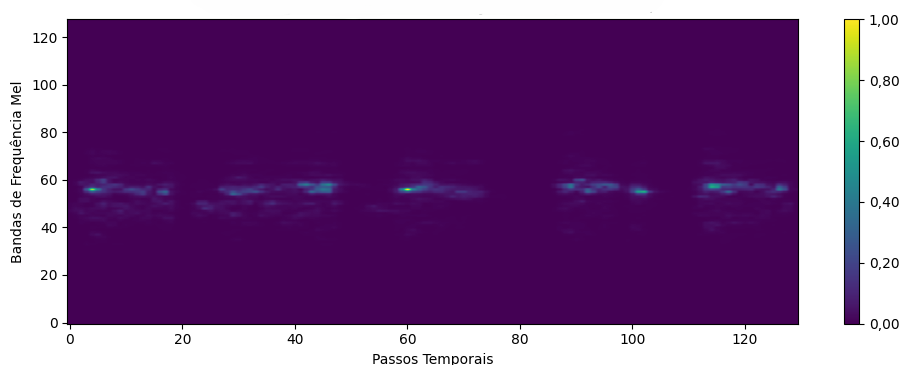


Figura 1. *Mel spectrogram* da vocalização de uma *Anodorhynchus hyacinthinus*

A conversão em *mel spectrograms* permite o uso de técnicas de visão computacional, como redes neurais convolucionais (CNNs), para reconhecimento de padrões acústicos, funcionando como uma ponte eficaz entre sinais acústicos e algoritmos de ML.

2.3. Machine Learning na Classificação de Sons de Aves

ML tem se estabelecido como uma abordagem eficaz para a análise automatizada de dados bioacústicos [Sharma et al. 2022]. Diversos algoritmos podem detectar e classificar vocalizações com alta precisão, mesmo em ambientes ruidosos [Ventura et al. 2024].

As CNNs são amplamente utilizadas na classificação de espectrogramas, capazes de aprender automaticamente padrões espaciais relevantes [Sharma et al. 2022]. A arquitetura típica inclui camadas convolucionais, de pooling e camadas densas, analisando espectrogramas para extrair características discriminativas. Outras técnicas eficazes incluem Random Forests (robusto a ruídos), Support Vector Machines (adequado para alta dimensionalidade), K-Nearest Neighbors (simples e eficiente) e Redes Neurais Recorrentes (para dados sequenciais) [Sharma et al. 2022].

A escolha do modelo apropriado depende das características do conjunto de dados, da qualidade dos espectrogramas e do objetivo do estudo. O uso de ML em bioacústica oferece vantagens como a automação do monitoramento e maior precisão, mas enfrenta desafios como ruído ambiental e escassez de dados rotulados para espécies raras.

3. Metodologia

Nesta seção, são abordados os procedimentos adotados para o desenvolvimento do estudo, incluindo a seleção das espécies-alvo, a construção e o pré-processamento do conjunto de dados, a extração de características dos sinais sonoros, a modelagem com algoritmos de ML e a estratégia utilizada para estimar o número de indivíduos a partir das vocalizações detectadas.

3.1. Espécies-alvo

As espécies de aves selecionadas para este estudo foram escolhidas com base em seu status de conservação, conforme os critérios estabelecidos pelo Livro Vermelho da Fauna Brasileira Ameaçada de Extinção (ICMBio, 2018), pela Lista Vermelha da União Internacional para a Conservação da Natureza (IUCN, 2024), além da orientação fornecida por [Nunes et al. 2021]. A seleção concentrou-se em espécies com ocorrência registrada no bioma Pantanal e que apresentam algum grau de ameaça à conservação.

Segundo o Livro Vermelho, havia, à época de sua publicação, 13 espécies de aves ameaçadas no Pantanal. No entanto, essa obra não delimita claramente as espécies pertencentes ao bioma, apresentando apenas mapas de ocorrência geral. Nesse contexto, o trabalho de [Nunes et al. 2021] foi fundamental para identificar as espécies efetivamente presentes na região.

Observou-se, entretanto, uma discrepância entre as duas fontes: o Livro Vermelho indica 13 espécies ameaçadas no bioma, enquanto [Nunes et al. 2021] aponta a existência de 25 espécies em risco. Para resolver essa divergência, adotou-se a lista de [Nunes et al. 2021], complementada por uma revisão atualizada do status de conservação nas fontes oficiais.

Após a consolidação das listas e análise dos critérios de conservação, foram identificadas 25 espécies potenciais para o estudo. Dessas, foram mantidas apenas aquelas listadas como ameaçadas em pelo menos um dos sistemas de classificação considerados (IUCN ou Livro Vermelho da Fauna Brasileira), resultando em um conjunto de espécies candidatas que inclui diferentes níveis de ameaça, desde Quase Ameaçada (NT) até Criticamente em Perigo (CR).

Considerando a necessidade de dados suficientes para a modelagem acústica, optou-se por selecionar as cinco espécies com maior número de gravações disponíveis na plataforma xeno-canto, visando garantir uma base representativa e estatisticamente robusta para o desenvolvimento dos modelos. A Tabela 1 apresenta as espécies finalmente escolhidas para a continuidade do trabalho.

3.2. Conjunto de dados

As amostras positivas consistem em gravações de vocalizações de cinco espécies de aves ameaçadas de extinção, extraídas da base pública *Xeno-Canto* [Xeno-canto Foundation and Naturalis Biodiversity Center 2024], uma das

Tabela 1. Espécies selecionadas para o estudo

Espécie	Nº de gravações
<i>Calidris canutus</i>	248
<i>Calidris pusilla</i>	121
<i>Anodorhynchus hyacinthinus</i>	86
<i>Harpia harpyja</i>	76
<i>Crax fasciolata</i>	72

principais plataformas colaborativas de compartilhamento de sons de aves do mundo. O sistema de classificação de qualidade varia da letra **A** (excelente qualidade) até **E** (qualidade muito baixa).

Foram utilizados 567 áudios das espécies: *Calidris canutus*, *Calidris pusilla*, *Anodorhynchus hyacinthinus*, *Harpia harpyja* e *Crax fasciolata*. A Tabela 2 apresenta a distribuição por espécie.

Tabela 2. Soma da duração total das gravações por espécie (em segundos)

Espécie	Nº de gravações	Duração Total (segundos)
<i>Anodorhynchus hyacinthinus</i>	84	3511,94
<i>Calidris canutus</i>	223	10131,10
<i>Calidris pusilla</i>	117	3628,41
<i>Crax fasciolata</i>	71	2524,35
<i>Harpia harpyja</i>	72	2356,28
Total	567	22152,08

As gravações apresentam duração de 0,41 a 780,75 segundos, com média de aproximadamente 39,7 segundos. A Tabela 3 apresenta as estatísticas descritivas. Já as amostras negativas, detalhadas na Tabela 4, são compostas por gravações de ambientes naturais e por vocalizações de espécies filogeneticamente próximas às espécies-alvo. As gravações incluem *Calidris melanotos*, *Calidris mauri*, *Anodorhynchus leari*, *Crax blumenbachii* e *Crax alector* (do *Xeno-Canto*), além de sons ambientes florestais da plataforma *Eco-sound*[Queensland University of Technology’s Ecoacoustics Research Group 2024].

Tabela 3. Estatísticas de duração por espécie (em segundos)

Espécie	Média	Desvio Padrão	25%	Mediana	75%
<i>Anodorhynchus hyacinthinus</i>	41,81	37,81	17,59	29,86	48,70
<i>Calidris canutus</i>	45,43	84,82	11,02	21,88	44,31
<i>Calidris pusilla</i>	31,01	44,28	7,13	18,31	40,17
<i>Crax fasciolata</i>	35,55	34,86	11,95	23,40	46,82
<i>Harpia harpyja</i>	32,73	27,26	14,40	24,32	42,67
Total	39,07	61,07	11,27	22,59	44,64

3.3. Pré-processamento

O pré-processamento é essencial para garantir a qualidade dos dados. Os arquivos de áudio são segmentados em janelas de 3 segundos para padronizar o comprimento das

Tabela 4. Amostras negativas

Item	Nº de gravações	Duração Total (segundos)
<i>Anodorhynchus leari</i>	10	193,76
<i>Calidris mauri</i>	10	660,58
<i>Calidris melanotos</i>	10	342,35
<i>Crax alector</i>	10	441,82
<i>Crax blumenbachii</i>	10	462,94
Sons ambientes florestais	75	1651,24
Total	125	3752,69

entradas. Cada segmento é classificado para identificar a presença de vocalizações da espécie-alvo, mantendo apenas os segmentos relevantes. Para amostras negativas, todos os segmentos são mantidos. O número de segmentos usados para cada grupo é descrito na Tabela 5.

Tabela 5. Quantidade de segmentos

Item	Nº de segmentos
<i>Anodorhynchus hyacinthinus</i>	888
<i>Calidris canutus</i>	1138
<i>Calidris pusilla</i>	1098
<i>Crax fasciolata</i>	439
<i>Harpia harpyja</i>	478
Amostras negativas	1248

Segmentos com duração inferior a 3 segundos são preenchidos com zeros (*zero-padding*). Para cada janela, é gerado um *mel spectrogram* com 128 bandas. Os espectrogramas são convertidos para escala de decibéis, normalizados para [0, 1] e reorganizados no formato (128, passos temporais, 1) para entrada nas redes convolucionais.

3.4. Arquitetura da Rede Neural Convolucional

A arquitetura, ilustrada pela 2, baseia-se em uma CNN projetada para classificação de áudio a partir de *mel spectrograms*. De acordo com [Sharma et al. 2022], CNN é uma técnica flexível e robusta, ainda que em conjuntos de dados pequenos. As configurações adotadas foram definidas com base em um processo iterativo de testes e adaptações, realizado com o objetivo de identificar a arquitetura que apresentasse um desempenho satisfatório frente aos dados utilizados. As entradas consistem em espectrogramas de trechos de 3 segundos, amostrados a 22.050 Hz e convertidos em 128 bandas mel-escaladas, organizados no formato (1, 128, T).

A rede é composta por três blocos convolucionais sucessivos:

- **Primeiro bloco:** Conv2D com 32 filtros (kernel 3×3), BatchNorm2D, ativação ReLU, *max pooling* 2×2 e Dropout de 20%.
- **Segundo bloco:** estrutura idêntica com 64 filtros e Dropout de 30%.
- **Terceiro bloco:** estrutura idêntica com 128 filtros e Dropout de 40%.

A classificação é realizada por duas camadas totalmente conectadas: a primeira com 128 unidades, ativação ReLU e Dropout de 50%; a segunda com um neurônio e ativação sigmoid para classificação binária.

O modelo foi compilado com otimizador Adam (taxa de aprendizado 0,001) e função de perda `binary_crossentropy`, utilizando *early stopping* e ajuste adaptativo da taxa de aprendizado.

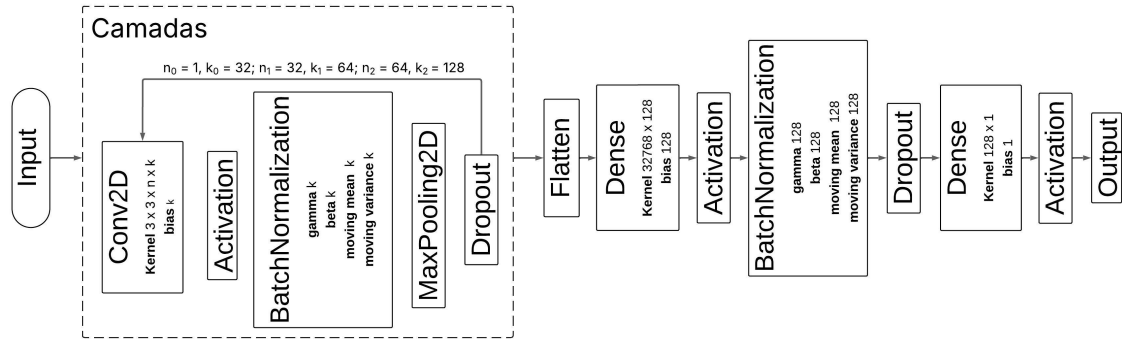


Figura 2. Arquitetura do modelo

3.5. Execução

3.5.1. Estratégia de divisão de dados, Aumento de dados e Ponderação de classes

Foi adotada divisão estratificada para preservar a proporção de classes: treinamento + validação (80%) e teste (20%), com subdivisão posterior em treinamento (64%) e validação (16%).

Para melhorar a generalização, especialmente para amostras positivas limitadas, foi implementado aumento de dados com *Frequency Masking* (2-4 máscaras por espectrograma), mascaramento temporal (5-20 quadros) e ruídos estocásticos. O procedimento foi aplicado com fator 2 às amostras positivas, atenuando o desbalanceamento. O método atua sobre *mel spectrograms*, aplicando uma abordagem mista semelhante à descrita por [Kumar et al. 2024].

Para lidar com o desbalanceamento, foi implementada ponderação de classes inversamente proporcional à frequência:

$$w_c = \frac{N}{2 \cdot n_c}$$

onde w_c é o peso da classe c , N é o total de amostras e n_c é o número de amostras da classe c .

3.5.2. Métricas de avaliação e Implementação

As métricas utilizadas para avaliar o desempenho dos modelos incluem a acurácia, que representa a proporção geral de predições corretas, e a precisão, que indica a capacidade do modelo de evitar falsos positivos. Também foi considerada a métrica de recall (ou

sensibilidade), que expressa a capacidade de identificar todas as ocorrências da espécie-alvo, além do F1-score, calculado como a média harmônica entre precisão e recall.

O desenvolvimento utilizou Python com PyTorch (implementação do modelo neural), Librosa (processamento de áudio), NumPy (operações numéricas), scikit-learn (divisão de dados e métricas) e Matplotlib (visualização). O treinamento implementou *early stopping* com paciência de 10 épocas e 50 épocas no máximo, preservando o *checkpoint* com menor perda de validação.

4. Resultados

Esta seção apresenta os resultados obtidos a partir do treinamento e avaliação do modelo de classificação binária. Os resultados estão segmentados por espécie, e ao final da seção são apresentados de forma conjunta.

4.1. *Anodorhynchus hyacinthinus*

O modelo treinado para a espécie *Anodorhynchus hyacinthinus* apresentou desempenho robusto no conjunto de teste, alcançando uma acurácia de 92%. O recall da classe positiva (espécie-alvo) foi de 0,94, indicando que o modelo conseguiu identificar a maioria das vocalizações verdadeiras da espécie. Esse resultado reflete uma boa sensibilidade, aspecto essencial em contextos de monitoramento de biodiversidade, onde a detecção eficaz da espécie-alvo é prioritária.

A precisão da classe positiva foi de 0,87, revelando um baixo índice de falsos positivos. O valor do F1-score para essa classe foi de 0,91, demonstrando um bom equilíbrio entre precisão e recall.

Durante o treinamento, o modelo atingiu sua melhor performance na época 28, com loss de 0,0355 e acurácia de 98,82% no conjunto de treinamento. O processo de treinamento foi interrompido automaticamente por meio de *early stopping*, ativado com base no parâmetro de paciência, o que evitou o sobreajuste ao não observar melhora nas métricas de validação ao longo de algumas épocas consecutivas. As métricas detalhadas de desempenho estão resumidas no relatório de classificação apresentado na tabela 6.

Tabela 6. Relatório de classificação para *Anodorhynchus hyacinthinus*



Classe	Precisão	Recall	F1-score	Suporte
0 (negativa)	0,95	0,91	0,93	257
1 (positiva)	0,87	0,94	0,91	178
Acurácia	0,92 (n = 435)			
Média ponderada	0,92	0,92	0,92	435

4.2. *Calidris canutus*

O modelo treinado para a espécie *Calidris canutus* apresentou desempenho satisfatório, atingindo uma acurácia de 93% no conjunto de teste. A precisão e o recall da classe positiva (espécie-alvo) foram de 0,91 e 0,93, respectivamente. Isso demonstra que o modelo teve um bom equilíbrio entre a capacidade de identificar corretamente as vocalizações da espécie e a limitação de falsos positivos.

O recall de 0,93 representa um desempenho consistente para a classe positiva, especialmente considerando que esse valor indica que a maioria das vocalizações da espécie foi corretamente detectada. O F1-score dessa classe foi de 0,92, reforçando o equilíbrio entre sensibilidade e precisão.

Durante o treinamento, o modelo alcançou sua melhor performance na época 46, número superior às demais espécies, com loss de 0,0480 e acurácia de 98,29% no conjunto de treinamento. O treinamento foi interrompido por *early stopping*, motivado pelo parâmetro de paciência configurado para evitar sobreajuste. As métricas detalhadas de desempenho estão resumidas no relatório de classificação apresentado na tabela 7.

Tabela 7. Relatório de classificação para *Calidris canutus*



Classe	Precisão	Recall	F1-score	Suporte
0 (negativa)	0,94	0,92	0,93	257
1 (positiva)	0,91	0,93	0,92	228
Acurácia	0,93 (n = 485)			
Média ponderada	0,93	0,93	0,93	485

4.3. *Calidris pusilla*

O modelo treinado para a espécie *Calidris pusilla* obteve uma acurácia de 88% no conjunto de teste, com desempenho assimétrico entre as classes. O recall da classe positiva (espécie-alvo) foi de 0,80, indicando que o modelo conseguiu identificar a maior parte das vocalizações da espécie. A precisão, por outro lado, foi de 0,93, mostrando que a maioria das detecções feitas como positivas realmente correspondia à espécie correta. O F1-score da classe positiva foi de 0,86, refletindo um bom equilíbrio entre sensibilidade e precisão.

O treinamento foi interrompido por *early stopping* na época 19, em função do parâmetro de paciência do modelo. Na melhor época, o modelo apresentou loss de 0,0548 e acurácia de 98,20% no conjunto de treinamento. As métricas detalhadas de desempenho estão resumidas no relatório de classificação apresentado na tabela 8.

Tabela 8. Relatório de classificação para *Calidris pusilla*



Classe	Precisão	Recall	F1-score	Suporte
0 (negativa)	0,85	0,95	0,90	257
1 (positiva)	0,93	0,80	0,86	220
Acurácia	0,88 (n = 477)			
Média ponderada	0,89	0,88	0,88	477

4.4. *Crax fasciolata*

O modelo treinado para a espécie *Crax fasciolata* apresentou desempenho notável, alcançando uma acurácia de 85% no conjunto de teste. O recall da classe positiva (espécie-alvo) foi de 0,70, o que demonstra que o modelo foi capaz de detectar corretamente a maioria das vocalizações reais da espécie. A precisão foi de 0,70, indicando que ainda houve uma quantidade considerável de falsos positivos. O F1-score da classe

positiva ficou em torno de 0,70, refletindo um equilíbrio razoável entre precisão e sensibilidade.

Durante o treinamento, o modelo obteve seu melhor desempenho na época 20, com loss de 0,04572 e acurácia de 97,47% no conjunto de treinamento. O processo foi interrompido automaticamente por *early stopping*, em função do critério de paciência configurado. As métricas detalhadas de desempenho estão resumidas no relatório de classificação apresentado na tabela 9.

Tabela 9. Relatório de classificação para *Crax fasciolata*



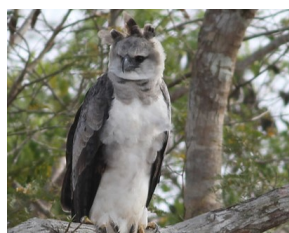
Classe	Precisão	Recall	F1-score	Suporte
0 (negativa)	0,90	0,89	0,90	257
1 (positiva)	0,70	0,70	0,70	88
Acurácia	0,85 (n = 345)			
Média ponderada	0,85	0,85	0,85	345

4.5. Harpia harpyja

O modelo treinado para a espécie Harpia harpyja apresentou excelente desempenho, alcançando uma acurácia de 96% no conjunto de teste. O recall da classe positiva (espécie-alvo) foi de 0,89, com uma precisão ainda mais elevada (0,97), resultando em um F1-score de 0,92 para essa classe. Isso demonstra que o modelo foi altamente eficaz em identificar corretamente as vocalizações reais da espécie.

O treinamento foi interrompido por *early stopping* na época 17, motivado pelo critério de paciência ($p = 5$). Nessa fase, o modelo atingiu loss de 0,0252 e acurácia de 99,37% no treinamento. As métricas detalhadas de desempenho estão resumidas no relatório de classificação apresentado na tabela 10.

Tabela 10. Relatório de classificação para *Harpia harpyja*



Classe	Precisão	Recall	F1-score	Suporte
0 (negativa)	0,96	0,99	0,97	257
1 (positiva)	0,97	0,89	0,92	96
Acurácia	0,98 (n = 353)			
Média ponderada	0,96	0,96	0,96	353

4.6. Comparativo entre espécies

A análise comparativa evidencia distintos comportamentos de desempenho entre os modelos treinados para cada espécie. Modelos como os de *Calidris pusilla* e *Harpia harpyja* apresentaram excelente equilíbrio entre precisão e recall, com valores de precisão de 0,97 e 0,93, destacando-se pela confiabilidade geral. Por outro lado, o modelo de *Calidris canutus* mostrou bom equilíbrio (precisão 0,91 e recall 0,93), enquanto o de *Crax fasciolata* apresentou o menor recall (0,70), o que é indesejado no contexto deste trabalho, onde falso positivos devem ser mínimos.

Já o modelo para *Anodorhynchus hyacinthinus* obteve precisão de 0,87 e recall de 0,94, demonstrando desempenho equilibrado, possuindo o maior recall dentre as espécies. Destaca-se o modelo para *Harpia harpyja*, que alcançou os melhores resultados em termos de acurácia geral (0,97), sugerindo alta capacidade de identificação correta das vocalizações desta espécie com certa ocorrência de falsos negativos (recall de 0,89). Os modelos para *Calidris canutus* e *Harpia harpyja* obtiveram os melhores F1-score (0,93).

A tabela 11 resume os principais indicadores para a classe positiva (espécie-alvo), permitindo uma avaliação direta entre as estratégias de detecção adotadas implicitamente por cada modelo.

Tabela 11. Resumo das métricas para a classe positiva por espécie-alvo

Espécie	Precisão	Recall	F1-score	Acurácia
<i>Anodorhynchus hyacinthinus</i>	0,87	0,94	0,91	0,92
<i>Calidris canutus</i>	0,91	0,93	0,92	0,93
<i>Calidris pusilla</i>	0,93	0,80	0,86	0,88
<i>Crax fasciolata</i>	0,70	0,70	0,70	0,85
<i>Harpia harpyja</i>	0,97	0,89	0,92	0,96

5. Considerações Finais

Este trabalho avaliou a viabilidade do uso de técnicas de *Deep Learning* e de técnicas de tratamento e aumento de dado aplicadas à bioacústica para identificar aves ameaçadas no Pantanal, desde a seleção criteriosa de espécies e dados até a implementação prática de modelos classificadores. A curadoria dos dados, ancorada em critérios da IUCN e do Livro Vermelho, e o uso de *mel spectrograms* com redes neurais convolucionais, permitiram mensurar o desempenho de modelos em diferentes contextos ecológicos.

Os resultados obtidos foram promissores, com destaque para *Harpia harpyja*, cuja acurácia (0,97) e F1-score (0,92) indicaram desempenho superior. *Anodorhynchus hyacinthinus* apresentou o maior recall (0,94), demonstrando alta sensibilidade, especialmente útil em contextos de preservação. Já *Calidris canutus* e *Calidris pusilla* obtiveram resultados satisfatórios, enquanto *Crax fasciolata* apresentou métricas mais modestas (0,70). A ênfase no recall mostrou-se essencial para reduzir falsos negativos, assegurando que eventos relevantes de vocalização não sejam negligenciados.

Apesar dos avanços, persistem desafios como a escassez de gravações rotuladas, a variabilidade dos sons em ambientes naturais e a limitação de generalização dos modelos. A superação dessas barreiras requer colaboração interinstitucional e investimento em repositórios bioacústicos mais abrangentes.

Em termos práticos, este estudo oferece uma base metodológica robusta para o uso de inteligência artificial em projetos de monitoramento da biodiversidade pantaneira, evidenciando sua aplicabilidade em estratégias de conservação sustentadas por critérios técnicos e ecológicos sólidos.

Desdobramentos futuros incluem o aprimoramento das abordagens modelares, com redes mais robustas e diversificação dos dados de treino. A adoção de técnicas auto-supervisionadas e novas formas de representação acústica também pode ampliar o desempenho. Por fim, propõe-se o desenvolvimento de um pipeline integrado, da coleta em

campo via PAM (*Passive Acoustic Monitoring* à classificação automatizada, viabilizando aplicações reais no monitoramento contínuo da avifauna ameaçada.

Referências

- de Oliveira, A. G., Ventura, T. M., Ganchev, T. D., Silva, L. N., Marques, M. I., and Schuchmann, K.-L. (2020). Speeding up training of automated bird recognizers by data reduction of audio features. *PeerJ*, 8:e8407.
- Instituto Chico Mendes de Conservação da Biodiversidade (2018). *Livro Vermelho da Fauna Brasileira Ameaçada de Extinção: Volume III – Aves*, volume 3 of *Livro Vermelho da Fauna Brasileira Ameaçada de Extinção*. ICMBio/MMA, Brasília, DF, 1 edition. 7 volumes. Conteúdo: v.1. Geral; v.2. Mamíferos; v.3. Aves; v.4. Répteis; v.5. Anfíbios; v.6. Peixes; v.7. Invertebrados (Pencard). Ilustrado.
- IUCN (2025). The iucn red list of threatened species. version 2025-1. <https://www.iucnredlist.org>. Acesso: 10 de abril de 2025.
- 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.
- Kumar, A. S., Schlosser, T., Kahl, S., and Kowerko, D. (2024). Improving learning-based birdsong classification by utilizing combined audio augmentation strategies. *Ecological Informatics*, 82:102699.
- Nunes, A. P., Posso, S. R., Frota, A. V. B. d., Vitorino, B. D., Laps, R. R., Donatelli, R. J., Straube, F. C., Pivatto, M. A. C., Oliveira, D. M. M. d., Carlos, B., and et al. (2021). Birds of the pantanal floodplains, brazil: historical data, diversity, and conservation. *Papéis Avulsos de Zoologia*, 61:e20216182.
- Queensland University of Technology’s Ecoacoustics Research Group (2024). Eco-sounds. Acesso: 10 de abril de 2025.
- Sharma, S., Sato, K., and Gautam, B. P. (2022). Bioacoustics monitoring of wildlife using artificial intelligence: A methodological literature review. In *2022 International Conference on Networking and Network Applications (NaNA)*. IEEE.
- Stowell, D. (2022). Computational bioacoustics with deep learning: a review and roadmap. *PeerJ*, 10:e13152.
- Ventura, T. M., Ganchev, T. D., Pérez-Granados, C., de Oliveira, A. G., de S. G. Pedroso, G., Marques, M. I., and and, K.-L. S. (2024). The importance of acoustic background modelling in cnn-based detection of the neotropical white-lored spinetail (aves, passeriformes, furnaridae). *Bioacoustics*, 33(2):103–121.
- Xeno-canto Foundation and Naturalis Biodiversity Center (2024). xeno-canto: Sharing wildlife sounds from around the world. Acesso: 10 de abril de 2025.
- Xie, J., Hu, K., Zhu, M., Yu, J., and Zhu, Q. (2019). Investigation of different cnn-based models for improved bird sound classification. *IEEE Access*, 7:175353–175361.