



## K. Lisa Yang Center for Conservation Bioacoustics

### TEMABio Pantanal – Tutorial de 25 de Junho de 2025

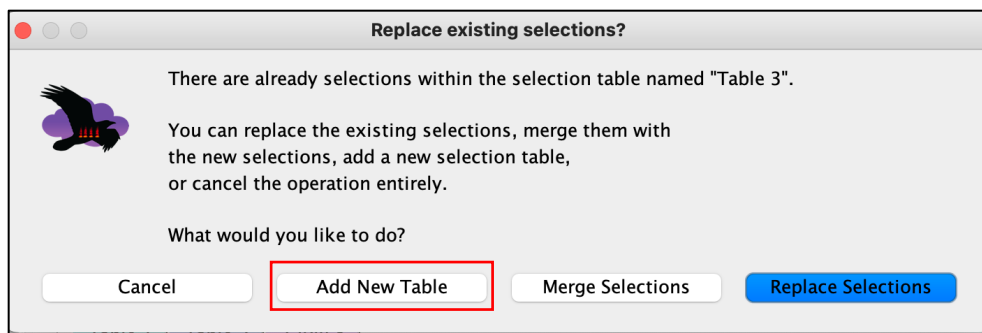
Atividade: Avaliando a performance de um modelo

**Objetivo:** Comparar predições de um modelo customizado BirdNET com dados anotados manualmente

Esse exercício utilizará a última versão do BirdNET Analyzer (2.0.1), a qual contém uma nova aba, gerada por um engenheiro de software voluntário, com uma série de ferramentas de auxílio a avaliação de performance de modelos.

Os dados foram gentilmente cedidos por Dena Clink e o exercício é baseado em um tutorial gerado por Ben Gottesman. Os dados totalizam 1 hora de áudios aleatoriamente selecionados a partir de 10 locais, todos anotados pelo pesquisador Ashraft Yusni. Os testes foram coletados em Jahoo, Malaysia.

1. Navegue até a pasta 3\_GreatHornbill\_Example\_Comparison\_Automated\_vs\_Manual e abra o arquivo de áudio R1064\_20220614\_100004\_Excerpt no Raven.
2. Adicione a tabela de anotações manuais gerada para esse áudio, R1064\_20220614\_100004.Table.1.selections.txt. Escute os sons do Calau.
3. Agora adicione a tabela R1064\_20220614\_100004.Table.1.selections.txt, selecionando a opção Add New Table.



4. Em nosso caso, iremos usar a mesma escala de detecção do BirdNET para assinalar uma detecção a um sinal anotado. Ou seja, a escala de 3segundos contendo ou não o Calau é a janela na qual estamos extraindo as métricas de performance. No entanto, não vamos fazer isso manualmente: a aba Evaluation permite automatizar o processo!



## K. Lisa Yang Center for Conservation Bioacoustics

5. Para tanto, basta que as detecções do BirdNET estejam compatíveis com os nomes das anotações (notem que nas pastas de predição e anotação, o nome dos áudios faz parte do nome dos arquivos).
6. Abra o BirdNET e navegue até a aba Evaluation.
7. Em Select annotation directory, selecione a pasta  
1\_GreatHornbill\_Annotation\_directory
8. Em Select prediction directory, selecione a pasta  
2\_GreatHornbill\_Prediction\_directory
9. Expanda todas as outras opções para explorar as possibilidades de configuração, mas mantenha os valores default.
10. Expanda a opção Parameters. Qual é o limiar (threshold)?
11. Clique em 'Calculate Metrics'. Qual é o valor de precisão e revocação?
12. Clique em 'Plot Metric for all Thresholds'. Veja o comportamento de Precision e Recall.
13. O que acontece com a precisão para diferentes valores de limiares (ou seja, a variação ao longo do eixo horizontal do gráfico)?
14. Faça o download dos dados em 'Download data table'. Abra o arquivo .csv, aplique um filtro e ordene os dados em ordem ascendente conforme valores da quinta coluna "Great Hornbill\_confidence".

filename	sample_id	start_time	end_time	Great Hornbill_confidence	Great Hornbill_label
R1063_20220	88	264	267	0.9999	1
R1063_20220	108	324	327	0.9998	1
R1063_20220	266	798	801	0.9998	1
R1063_20220	275	825	828	0.9998	1
R1063_20220	99	297	300	0.9996	1
R1063_20220	316	948	951	0.9996	1
R1063_20220	265	795	798	0.9995	1
R1063_20220	503	1509	1512	0.9994	1
R1063_20220	104	312	315	0.9992	1
R1063_20220	409	1227	1230	0.9991	1
R1063_20220	300	900	903	0.9991	1
R1063_20220	1036	3108	3111	0.999	1
R1063_20220	107	321	324	0.9987	1
R1063_20220	473	1419	1422	0.9987	1
R1063_20220	97	291	294	0.9985	1
R1062_20220	81	243	246	0.9984	1
R1063_20220	320	960	963	0.9983	0
R1063_20220	273	819	822	0.9983	1
R1063_20220	522	1566	1569	0.9982	1

12. Vamos supor que você escolha um limiar de 0.7, que segundo o gráfico, tem uma precisão superior a 0.9. Ou seja, há muita certeza de que as detecções acima desse valor realmente estão corretas. No entanto, note a quantidade de valores abaixo de 0.7 que



## K. Lisa Yang Center for Conservation Bioacoustics

possuem a sexta coluna 'Great Hornbill\_annotation' com valor igual a 1. Todas essas predições estão sendo ignoradas.

Isso quer dizer que um valor alto de precisão não necessariamente indica que seu modelo está detectando tudo. Apenas que ele está muito certo do que está detectando. Quando você olha o recall no limar de 0.7, a métrica é muito baixa.

13. Qual limiar você escolheria para utilizar em um conjunto de dados, considerando os resultados encontrados?