**MACHINE LEARNING E CONSERVAÇÃO: PREVISÃO DE CATEGORIAS DE AMEAÇA EM ELASMOBRÂNQUIOS COM DADOS DEFICIENTES**

Ingrid Vasconcellos Bunholi¹; Leonardo Fonseca Reginato2; Marcela Alvarenga3; Rusty A. Gonser4

1Pós-Graduanda em Marine Sciences, University of Texas at Austin Marine Science insitute - UTMSI, Texas, Estados Unidos

2Pós-Graduando em Engenharia de Petróleo, Universidade de São Paulo – USP, Brasil

3Pós-Graduanda em Biodiversidade, Genética e Evolução, Centro de Investigação em Biodiversidade e Recursos Genéticos – CIBIO, Universidade do Porto, Portugal

4Professor no Biology Department, Indiana State University – ISU, Indiana, Estados Unidos

A subclasse Elasmobranchii (tubarões e raias) sobreviveu a diversos eventos de extinção em massa, confirmando sua alta resiliência frente a impactos ambientais. No entanto, esse grupo tem sofrido grandes declínios populacionais devido as atividades antrópicas e, consequentemente, 37% das espécies estão atualmente classificadas em algum grupo de ameaça de extinção (IUCN, 2022). Além disso, 16% estão categorizadas como “Dados Deficientes (DD)” pois não há informações suficientes para classificá-las dentro de uma categoria específica, o que acaba sendo uma grande lacuna para planos de manejo e conservação. Nesse sentido, o presente trabalho teve como objetivo aplicar ferramentas de aprendizado de máquina (do inglês *Machine Learning* – ML) para prever a possível categoria dessas espécies DD. A base de dados para a classificação foi obtida nos sites: (i) IUCN e (ii) Fishbase.com que contém variáveis biológicas, geográficas e ecológicas, consideradas importantes para a categorização de estados de ameaça. Um total de 1038 espécies e 18 variáveis foram obtidas. Através dos dados conhecidos, selecionamos a categoria de modelo supervisionado. Dentre os modelos testados, o algoritmo de floresta aleatória (do inglês *Random Forest*) foi o que obteve o melhor resultado médio de acurácia (~62.1%), sendo o selecionado para as previsões. Através deste modelo, as 114 espécies DD foram classificadas como: Menos preocupante (104 espécies), Vulnerável (oito espécies) e Ameaçada (duas espécies). Com isso, através do procedimento de coleta e tratamento de dados, combinado ao algoritmo de ML, foi obtido um indicativo de que 10 espécies DD podem pertencer a categorias de ameaça (Vulnerável e Ameaçada). Embora a maioria das espécies DD tenham sido classificadas em uma categoria de não-ameaça (Menos preocupante), deve-se considerar que espécies com ausência de dados devem ser prioritárias em pesquisas e planos de conservação devido a incerteza do seu real status. O uso de ferramentas de ML suporta ações imediatas para espécies classificadas dentro de categorias de ameaça, porém, não descarta o alerta para as demais. Além disso, das 10 espécies classificadas dentro de categorias de ameaça, 7 delas são espécies de raias, o que indica que estudos nesse grupo devem ser priorizados. Portanto, os resultados obtidos mostram a potência do uso de ferramentas de ciências de dados no auxílio a conservação, e que a classe DD pode estar negligenciando grandes declínios populacionais. Em conclusão, ferramentas e metodologias de ciência de dados possuem uma aplicabilidade abrangente na conservação de elasmobrânquios e devem ser consideradas como uma alternativa ou complemento na obtenção de informações, como por exemplo, para espécies de difícil acesso e espécies endêmicas de regiões com menos financiamento, visto ser uma alternativa acessível.

Palavras-chave: conservação; ciência de dados; IUCN