

Aplicação de técnicas de *Co-Teaching* para Limpeza de Dados e Treinamento Robusto de Redes Neurais: um Estudo de Caso de Modelagem de Distribuição de Espécies

Renato O. Miyaji¹, Felipe V. de Almeida¹, Pedro L. P. Corrêa¹

¹Escola Politécnica – Universidade de São Paulo (USP)

{re.miyaji, felipe.valencia.almeida, pedro.correa}@usp.br

Abstract. *In Classification tasks, when uncertainties are associated with the labels of the dataset, the training process of Neural Networks can be jeopardized. In this context, algorithmic techniques for data cleaning are effective. Thus, in this work, Co-Teaching techniques were applied, along with methods to address class imbalance, to identify instances with lower confidence in their labels, clean the data, and enable robust training of Neural Networks. These techniques rely on the joint training of multiple models and were applied in a Case Study of Species Distribution Modeling in the Amazon Basin region with a Neural Networks, achieving a 11.8% increase in ROC-AUC (increase from 75.8% to 84.5%).*

Resumo. *Em tarefas de Classificação, quando existem incertezas associadas aos rótulos do conjunto de dados, o ajuste de Redes Neurais pode ser comprometido. Nesse contexto, técnicas algorítmicas para limpeza de dados são efetivas. Assim, neste trabalho foram aplicadas técnicas de Co-Teaching, em conjunto com métodos para tratar o desbalanceamento entre classes, para identificar instâncias com menor confiança em seus rótulos, limpar os dados e permitir o treinamento robusto de Redes Neurais. Essas técnicas se baseiam no treinamento conjunto de múltiplos modelos e foram aplicadas em um Estudo de Caso de Modelagem de Distribuição de Espécies na região da Bacia Amazônica com um modelo de Redes Neurais, atingindo um incremento de ROC-AUC de 11,8% (aumento de 75,8% para 84,5%).*

1. Introdução

Em tarefas de classificação, é comum existirem inconsistências ou erros nos rótulos da variável resposta quando são utilizados conjuntos de dados reais. O ajuste de classificadores pode ser comprometido caso os erros e as inconsistências aconteçam com frequência. Normalmente, duas abordagens são utilizadas para melhorar tarefas de classificação. A primeira delas é da Inteligência Artificial centrada em Modelos (*Model-Centric Artificial Intelligence*) [Hamid 2022], na qual o processo de treinamento dos modelos é modificado, de modo a se tornarem mais robustos.

Por outro lado, outra abordagem possível é da Inteligência Artificial centrada em Dados (*Data-Centric Artificial Intelligence*) [Hamid 2022]. Por meio dela, as modificações não são concentradas nos modelos preditivos em si, mas no conjunto de dados utilizado para seu treinamento. Essas técnicas podem realizar de fato mudanças nos dados através de operações, como filtragens ou transformações, ou podem extrair informações

sobre o conjunto de dados e utilizá-las para alterar o processo de treinamento do modelo preditivo.

Quando são utilizados modelos preditivos baseados em Redes Neurais Artificiais, esses desafios podem se tornar ainda maiores, por conta de sua capacidade de memorização e consequente risco a sobreajuste [Han et al. 2018]. Assim, na literatura foram propostos diferentes métodos para viabilizar sua aplicação em conjuntos de dados nos quais existem incertezas associadas aos rótulos da variável resposta. Uma delas é a *Co-Teaching*, sendo baseada no treinamento de duas Redes Neurais simultaneamente. A cada lote do conjunto de dados (*Mini-Batch*) durante o treinamento, as redes se comunicam e trocam informações para identificar em conjunto amostras que possivelmente possuam rótulos corretos. Em seguida, cada uma realiza o processo de retropropagação (*Back-propagation*) com base nos rótulos que passaram pelo processo de avaliação por pares, atualizando seus pesos [Han et al. 2018].

Na área da Ecologia, a Modelagem de Distribuição de Espécies é utilizada para análises relacionadas à biodiversidade, em especial para avaliar os efeitos das variáveis ambientais na ocorrência de espécies e em seu nicho ecológico, isto é, as faixas de valores das variáveis que descrevem as condições ideais para o habitat das espécies analisadas. Nos últimos anos, existe uma grande tendência de utilizar técnicas de Aprendizado de Máquina para essa tarefa, por conta do seu desempenho superior a outras abordagens. Em especial, são aplicados modelos de Classificação binários, nos quais a classe positiva representa a ocorrência da espécie e a negativa, a sua ausência.

Uma dificuldade na utilização de Classificadores para a Modelagem de Distribuição de Espécies é relacionada aos rótulos da variável resposta durante a construção do conjunto de dados. Normalmente, são disponibilizados apenas dados com a presença da espécie, ou seja, a classe positiva (*Presence-only Data*). Isso ocorre, pois é o formato mais comum de dados de ocorrência de espécies, além da afirmação da ausência da espécie em determinada localização ser uma tarefa difícil de se realizar de maneira massificada. Dado este contexto, geralmente os conjuntos de dados utilizados para Modelagem de Distribuição de Espécies apresentam grandes incertezas e possíveis erros de rotulagem [Beery et al. 2021]. As incertezas a respeito dos rótulos nos conjuntos de dados de Modelagem de Distribuição de Espécies podem levar a resultados incorretos [Martin et al. 2005]. Assim, a aplicação de técnicas baseadas no conceito de *Co-Teaching* pode ser especialmente útil [Vos et al. 2023].

Dessa forma, o objetivo deste trabalho foi de aplicar técnicas baseadas no conceito de *Co-Teaching* para viabilizar o ajuste de modelos de Classificação de Aprendizado de Máquina que utilizam Redes Neurais, através da limpeza de dados com rótulos incertos e do treinamento robusto dos Classificadores. Para possibilitar a comparação e validação dessas técnicas, foi realizado um estudo de caso, no qual os modelos de Classificação foram utilizados para a tarefa de Modelagem de Distribuição de Espécies.

O restante deste documento está dividido em quatro seções principais. Na Seção 2, serão apresentados os trabalhos relacionados. Na Seção 3, é apresentada a metodologia utilizada. Na Seção 4, os resultados obtidos são discutidos. Por fim, são destacadas as conclusões.

2. Trabalhos Relacionados

Um desafio relevante na Ciência de Dados é em relação ao tratamento de dados inconsistentes ou incorretos. Esses podem comprometer o desenvolvimento de análises e a construção de modelos preditivos. Assim, na literatura foram propostas diferentes técnicas para o tratamento de incertezas nos dados.

Nos trabalhos de [Malach and Shalev-Shwartz 2017] e [Jiang et al. 2018] foram desenvolvidos métodos baseados no conceito de Ensino e Aprendizagem. Esses possuem como base principal o processo iterativo de identificação de instâncias com rótulos incorretos e posterior atualização da Rede Neural. Conforme as iterações do processo são feitas, obtém-se um conjunto de dados de treinamento cada vez menos incerto e, conseqüentemente, um modelo com melhor capacidade preditiva.

No método *Decoupling*, duas Redes Neurais são treinadas de forma simultânea com os mesmos dados. A atualização de seus pesos é feita a partir das instâncias nas quais a previsão delas não coincide. Entretanto, ainda existe uma dificuldade em tratar dos rótulos incertos de forma mais explícita [Malach and Shalev-Shwartz 2017]. O método *MentorNet* também utiliza duas Redes Neurais. Porém, uma delas é treinada primeiro e é utilizada para a seleção das instâncias com rótulos corretos, chamada de *Mentor*. A partir deles, é realizado o treinamento da segunda Rede [Jiang et al. 2018]. No entanto, existem limitações e riscos relacionados ao viés de seleção de amostra [Han et al. 2018].

Para mitigar os riscos dos demais métodos, [Han et al. 2018] propõe a técnica *Co-Teaching*. Ela também faz o uso de duas Redes Neurais que são treinadas simultaneamente, porém a medida que o treinamento é realizado em cada lote do conjunto de dados, é realizada uma comunicação entre elas sobre as instâncias com menor perda da função de custo. A partir disso, cada Rede atualiza seus pesos com a retropropagação. Por meio desse cruzamento de informações entre as Redes durante o treinamento, com o método *Co-Teaching*, a taxa de erro pode ser reduzida nas duas Redes de forma mútua [Han et al. 2018]. Uma limitação do método *Co-Teaching* é a necessidade de se saber, a priori, o nível de incerteza nos dados. Para superá-la, no trabalho de [Vos et al. 2023] é proposta a técnica *Stochastic Co-Teaching*, que utiliza de estocasticidade para selecionar ou rejeitar instâncias de treinamento.

Para a tarefa do estudo de caso, a Modelagem de Distribuição de Espécies, na literatura foram aplicadas diferentes técnicas para tratar as incertezas associadas aos rótulos, porém sem utilizar classificadores de Redes Neurais. Dessa forma, levando em consideração os resultados da revisão bibliográfica, nota-se que a aplicação de técnicas de *Co-Teaching* para Modelagem de Distribuição de Espécies é inédita na literatura.

3. Metodologia

3.1. *Co-Teaching*

Os conceitos de *Co-Teaching* aplicados a Redes Neurais Artificiais se diferenciam em relação a outras técnicas, como *Decoupling* e *MentorNet*, pela forma como as duas Redes Neurais se comunicam durante seu processo de treinamento, como é apresentado na Figura 1.

Assim como no caso do *Decoupling*, no *Co-Teaching* as duas Redes Neurais também são treinadas simultaneamente. A principal diferença entre os métodos é que a comu-

nicação entre as Redes ocorre de maneira cruzada a cada lote do conjunto de dados. Além disso, para selecionar as instâncias que serão utilizadas para atualizar os pesos de cada uma das Redes, é utilizado o critério de menor perda da função de custo [Han et al. 2018].

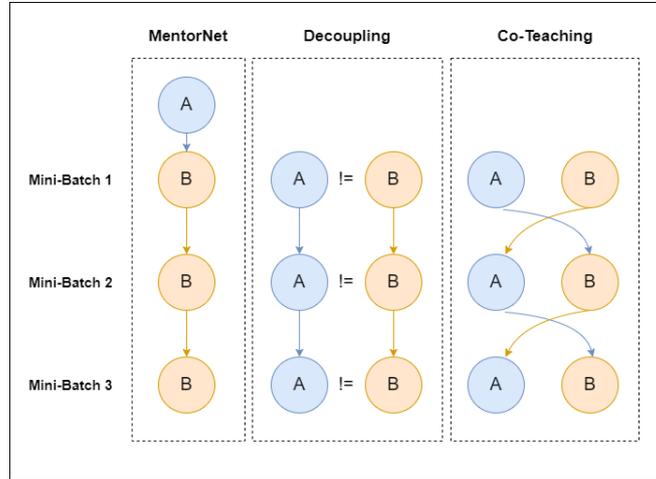


Figura 1. Comparação entre métodos *MentorNet*, *Decoupling* e *Co-Teaching*. Adaptado de [Han et al. 2018]

O método *Co-Teaching* é descrito no Algoritmo 1. Como se tratam de duas Redes Neurais Artificiais sendo treinadas simultaneamente, existem seus parâmetros w_1 e w_2 . A cada iteração, é construído um lote do conjunto de dados \tilde{D} . Para cada lote, é feita a propagação direta em ambas as Redes Neurais, obtendo as instâncias com menor perda da função de custo para cada Rede \tilde{D}_1 e \tilde{D}_2 . As instâncias selecionadas para compor \tilde{D}_1 e \tilde{D}_2 satisfazem o critério definido pela taxa de esquecimento p . Então, os parâmetros w_1 e w_2 de cada Rede Neural são atualizados por meio da Regra de Atualização U e as instâncias selecionadas pela Rede Neural par \tilde{D}_1 ou \tilde{D}_2 [Han et al. 2018].

A definição do parâmetro da taxa de esquecimento p varia a depender da abordagem na literatura. Na proposta original de [Han et al. 2018], a quantidade de instâncias selecionadas com menor perda da função de custo é definida dinamicamente a cada época T . O autor sugere que ela seja definida através de $R(T) = 1 - \min\{\frac{T}{T_k}\tau, \tau\}$, onde o parâmetro τ é definido a priori, levando em consideração o nível de incertezas presente no conjunto de dados D .

Entretanto, a quantificação do parâmetro τ pode ser complexa. Por isso, no trabalho de [Vos et al. 2023] é proposto o método *Stochastic Co-Teaching* que não requer essa definição inicial. Nele, é utilizada um abordagem estocástica na qual a cada iteração, o parâmetro p é definido de maneira aleatória a cada iteração a partir de uma distribuição de probabilidade Beta, que possui valor entre 0 e 1, sendo controlada pelos parâmetros α e β . Caso os parâmetros sejam iguais, a distribuição é simétrica, mas pode se tornar bimodal, uniforme ou pode possuir assimetria positiva ou negativa, conforme os parâmetros variem [Vos et al. 2023].

3.2. Estudo de Caso

Para o Estudo de Caso foi realizada uma Modelagem de Distribuição de Espécies. Trata-se de uma tarefa na qual a aplicação de técnicas *Co-Teaching* é especialmente útil, pois existem incertezas associadas aos rótulos da variável resposta [Martin et al. 2005].

Algorithm 1: Co-Teaching

Input: Modelos w_1 e w_2 , Taxa de Esquecimento p , Época T_k e T_{max} na iteração N_{max} e Regra de Atualização U

Data : Conjunto de Dados de Treinamento D

```

1 for  $T = 1, 2$  até  $T_{max}$  do
2   Aleatorização de Conjunto de Dados de Treinamento  $D$ ;
3   for  $N = 1$  até  $N_{max}$  do
4     Obter Mini-Batch  $\tilde{D}$  de  $D$ ;
5     Obter  $\tilde{D}_1 = \{(x_i, y_i) : p < P(x_i|w_1, y_i)\}$ ;
6     Obter  $\tilde{D}_2 = \{(x_i, y_i) : p < P(x_i|w_2, y_i)\}$ ;
7     Atualizar  $w_1 \leftarrow U(w_1, \tilde{D}_2)$ ;
8     Atualizar  $w_2 \leftarrow U(w_2, \tilde{D}_1)$ ;
9   end
10 end

```

O conjunto de dados utilizado para a Modelagem de Distribuição de Espécies foi obtido a partir de duas fontes principais. A primeira delas é relacionada às variáveis atmosféricas e de aerossóis que descrevem as condições ambientais, sendo elas: a temperatura, as concentrações de ozônio (O_3), monóxido de carbono (CO), óxidos de nitrogênio (NO_x), metano (CH_4), dióxido de carbono (CO_2), isopreno, acetona, a contagem numérica de partículas e a fração volumétrica de água (H_2O). Elas foram disponibilizadas por [Miyaji et al. 2021] e são resultantes de um processo de interpolação espacial a partir de coletas feitas por aeronaves na região analisada.

O Estudo de Caso se refere à região da Floresta Amazônica, nas proximidades de Manaus (AM). Trata-se de uma área que compreende um laboratório ideal para analisar a influência de ações antrópicas no clima e nos ecossistemas terrestres em contextos de florestas tropicais. Para a coleta dos dados referentes às condições ambientais, foram realizados voos de baixa altitude durante o projeto *Green Ocean Amazon 2014/15*, que foi feito pelo *Atmospheric Radiation Measurement (ARM)*, órgão do Departamento de Energia dos Estados Unidos da América, em parceria com instituições brasileiras [Martin et al. 2016].

A segunda fonte de dados utilizada foi para obter os dados de ocorrência das espécies. Esses foram obtidos a partir dos repositórios do *Global Biodiversity Information Facility (GBIF)* [GBIF 2024] e do Instituto Chico Mendes de Conservação da Biodiversidade (ICMBio) [ICMBio 2024]. Os dados foram extraídos para a mesma região e durante o mesmo período em que ocorreu a coleta do projeto GoAmazon 2014/15. Realizando o cruzamento entre as fontes de dados, foi possível gerar o conjunto de dados para realizar a Modelagem de Distribuição de Espécies.

Como ocorre comumente em tarefas de Modelagem de Distribuição de Espécies, a frequência de ocorrências das espécies analisadas não é muito elevada, o que pode tornar complexa a aplicação de modelos de Aprendizado de Máquina. Em específico para o

Estudo de Caso, a espécie com maior quantidade de ocorrências foi a *Coragyps atratus* (urubu de cabeça preta) com 54 registros distintos, que foi selecionada para análise por esse fator.

O conjunto de dados bioclimáticos utilizado era composto de dez variáveis climáticas e uma variável resposta: a ocorrência de *Coragyps atratus*. A região de abrangência dos dados é compreendida entre $-3,632^\circ$ e $-2,813^\circ$ de latitude e entre $-60,821^\circ$ e $-59,937^\circ$ de longitude, no centro da Bacia Amazônica. O período de coleta dos dados é entre agosto e outubro de 2014, durante a estação seca, sendo que os dados representam as médias deste período. O cruzamento entre os dados foi realizado considerando as chaves de coordenadas geográficas. Foi necessário adequar a resolução espacial para realizar o cruzamento, adotando a menor granularidade ($0,001^\circ$). Assim, o conjunto de dados gerado tinha dimensões de 185355 linhas e 11 colunas, sendo 7,6% das instâncias da classe positiva e o restante da negativa.

Existia um desbalanceamento natural entre as classes da variável resposta. Por isso, aplicou-se uma técnica de reamostragem no conjunto de dados para aumentar a frequência da classe minoritária: a *Synthetic Minority Oversampling Technique* (SMOTE). No conjunto de dados reamostrado, obteve-se uma proporção de 1:3 entre instâncias positivas e negativas. Para mitigar a introdução de vieses com a criação das amostras sintéticas, foi adotado um valor de vizinhança baixo de 2, de modo a se evitar grandes extrapolações. Não foram utilizados métodos de *Undersampling*, pois esses poderiam acarretar na perda de informação com a eliminação de uma grande quantidade de dados acerca da classe majoritária.

A arquitetura da Rede Neural Artificial a ser utilizada foi selecionada com base nas aplicadas em trabalhos com escopos semelhantes [Rew et al. 2021]. Assim, ela era composta de quatro camadas intermediárias com função de ativação *ReLU* com 16, 32, 64 e 128 neurônios respectivamente e camadas de *Drop Out* entre elas com taxas de 0.2, 0.3 e 0.5. Na camada de saída, a função de ativação era Sigmoid. Utilizou-se o otimizador *Stochastic Gradient Descent* (SGD) e inicializador *Glorot normal*.

Foram aplicados os métodos de *Co-Teaching* e *Stochastic Co-Teaching* durante o treinamento. Foram testados diferentes parâmetros de treinamento, como a taxa de aprendizado. Para determinação da melhor configuração, foi realizada uma Validação Cruzada com *5-Fold*. Com as métricas, foi realizado um Teste t Pareado a nível de significância de 5% a fim de se determinar a relevância estatística da diferença de desempenho. As métricas de avaliação dos modelos foram a Acurácia e a *Area Under the Receiver Operating characteristic Curve* (ROC-AUC), que foram avaliadas em um conjunto de dados de *Hold-Out* que representa 30 % dos dados originais.

4. Resultados e Discussões

Como referência de desempenho, inicialmente foi treinada a Rede Neural sem aplicação de técnicas *Co-Teaching*. O treinamento foi realizado com 300 épocas, tamanho de lote com 200 amostras, com a implementação de um mecanismo de *Early Stopping* e com uma taxa de aprendizado adaptativa que era reduzida gradualmente, iniciando com o valor de 0,1 e finalizando com 0,001. Esses hiper parâmetros foram obtidos a partir de um processo de Validação Cruzada. Avaliado sobre o conjunto de dados de *Hold-Out*, obteve-se uma Acurácia de 98,7 % e ROC-AUC de 75,8 %.

Aplicando o método *Co-Teaching* e utilizando a mesma arquitetura e os mesmos parâmetros de treinamento, foram avaliados diferentes valores de Taxa de Esquecimento p para permitir o melhor treinamento dos modelos. Utilizando um $p = 20\%$, obteve-se uma Acurácia de 99,3 % e ROC-AUC de 84,5 %. Já para aplicar o método *Stochastic Co-Teaching*, além de adotar os parâmetros de treinamento e a amostragem estratificada citados, também foram avaliados os parâmetros α e β da distribuição beta. O melhor desempenho foi obtido com $\alpha = 2$ e $\beta = 5$, com os quais a distribuição possui média de 28,6 % e moda de 20,0 %. Com essa configuração, obteve-se uma Acurácia de 99,1 % e ROC-AUC de 80,8 %.

Método	Acurácia	ROC-AUC
DNN	98,7% \pm 1,3%	75,8% \pm 2,1%
DNN-CT	99,3% \pm 1,1%	84,5% \pm 2,3%
DNN-SCT	99,1% \pm 1,5%	80,8% \pm 2,4%

Tabela 1. Comparação das métricas de classificação para modelos *Deep Neural Network* (DNN), *Co-Teaching* (DNN-CT) e *Stochastic Co-Teaching* (DNN-SCT)

Como pode ser observado na Tabela 1, com a aplicação dos métodos *Co-Teaching* e *Stochastic Co-Teaching* foi possível realizar uma limpeza no conjunto de dados com rótulos incertos e promover um treinamento mais robusto do Classificador. Desse modo, foram obtidas métricas de classificação superiores à referência de desempenho com relevância estatística para o incremento em ROC-AUC. Para a utilização dessas técnicas, especialmente a *Co-Teaching*, é necessário a priori definir um nível de incerteza para ser utilizado como Taxa de Esquecimento, que influencia diretamente na eficácia do método. Já com a técnica *Stochastic Co-Teaching* isso não é um requisito, apesar de ser necessário avaliar diferentes distribuições para determinar a mais adequada.

5. Conclusão e Trabalhos Futuros

Com este trabalho, foi possível avaliar a viabilidade de aplicação de técnicas de *Co-Teaching* para tarefas de Classificação com modelos de Aprendizado de Máquina baseados em Redes Neurais Artificiais em contextos nos quais existem incertezas associadas aos rótulos da variável resposta. Isso foi feito através de um Estudo de Caso para a tarefa de Modelagem de Distribuição de Espécies para a espécie *Coragyps atratus* a partir de variáveis atmosféricas e de aerossóis na região da Floresta Amazônica. Foi utilizado um conjunto de dados balanceado com a técnica *Synthetic Minority Oversampling Technique*, sobre o qual foi possível aplicar os métodos *Co-Teaching* e *Stochastic Co-Teaching*. Assim, obteve-se um incremento de desempenho para 84,5 % de ROC-AUC para o método *Co-Teaching* e 80,8 % para o *Stochastic Co-Teaching*, superando a referência de desempenho de 75,8 %.

A partir dos resultados obtidos, para trabalhos futuros sugere-se que os métodos baseados em *Co-Teaching* possam ser avaliados em experimentos com conjuntos de dados com espécies distintas (diferentes balanceamentos) e com arquiteturas mais complexas de Redes Neurais, como *Ensemble* ou Residuais.

Agradecimentos

O presente trabalho foi realizado com apoio da Coordenação de Aperfeiçoamento de Pessoal de Nível Superior – Brasil (CAPES) – Código de Financiamento 001. Ele foi possível devido aos Projetos Temáticos da FAPESP "Ciclos de vida e nuvens de aerossóis na Amazônia"(2017/ 17047-0) e "Research Centre for Greenhouse Gas Innovation - RCG2I"(2020/15230-5).

Referências

- Beery, S., Cole, E., Parker, J., Perona, P., and Winner, K. (2021). Species distribution modeling for machine learning practitioners: A review. In *Proceedings of ACM SIGCAS Conference on Computing and Sustainable Societies (COMPASS) 2021*.
- GBIF (2024). Gbif | global biodiversity information facility. <https://www.gbif.org/>. Acesso em: 2024-05-14.
- Hamid, O. H. (2022). From model-centric to data-centric ai: A paradigm shift or rather a complementary approach? In *Proceedings of 2022 8th International Conference on Information Technology Trends (ITT)*, pages 45–54. IEE.
- Han, B., Yao, Q., Yu, X., Niu, G., Xu, M., Hu, W., Tsang, I., and Sugiyama, M. (2018). Co-teaching: Robust training of deep neural networks with extremely noisy labels. In *Proceeding of the 32nd Conference on Neural Information Processing Systems (NeurIPS 2018)*.
- ICMBio (2024). Portal da biodiversidade do instituto chico mendes de conservação da biodiversidade. <https://portaldabiodiversidade.icmbio.gov.br/portal/>. Acesso em: 2024-05-14.
- Jiang, L., Zhou, Z., Leung, T., Li, L., and Fei-Fei, L. (2018). Mentornet: Learning data-driven curriculum for very deep neural networks on corrupted labels. In *Proceeding of the International Conference on Machine Learning (ICML 2018)*.
- Malach, E. and Shalev-Shwartz, S. (2017). Decoupling “when to update” from “how to update”. In *Proceeding of the Conference on Neural Information Processing Systems (NIPS 2017)*.
- Martin, S. T., Artaxo, P., Machado, L. A. T., Manzi, A. O., Souza, R. A. F. d., Schumacher, C., Wang, J., Andreae, M. O., Barbosa, H., Fan, J., et al. (2016). Introduction: observations and modeling of the green ocean amazon (goamazon2014/5). *Atmospheric Chemistry and Physics*, 16(8):4785–4797.
- Martin, T. G., Kuhnert, P. M., Mengersen, K., and Possingham, H. P. (2005). The power of expert opinion in ecological models using bayesian methods: Impact of grazing on birds. *Ecological Applications*, 15:266–280.
- Miyaji, R. O., Almeida, F. V., Bauer, L. O., Ferrari, V., Corrêa, P. L. P., Rizzo, L. V., and Prakash, G. (2021). Spatial interpolation of air pollutant and meteorological variables in central amazonia. *Data*, 6(12).
- Rew, J., Cho, U., and Hwang, E. (2021). A robust prediction model for species distribution using bagging ensembles with deep neural networks. *Remote Sensing*, 13(8):1945.
- Vos, B., Jansen, G., and Isgum, I. (2023). Stochastic co-teaching for training neural networks with unknown levels of label noise. *Scientific Reports*, 13(16875).