

Classificação de Solos com CNNs: Um Estudo Comparativo entre AlexNet e ResNet-18

Renato da Silva Coêlho

Instituto de Ciências Exatas e Tecnológicas

Universidade Federal de Viçosa - Campus Rio Paranaíba

Rio Paranaíba, Brasil

Matrícula: 8181

renato.d.coelho@ufv.br

Abstract—Soil classification is a task of paramount importance for agricultural management, yet it remains a labor-intensive manual process. Convolutional Neural Networks (CNNs) have emerged as a promising tool for automating this procedure, although their performance can be hindered when applied to small, imbalanced datasets. This paper compares the efficacy of two established architectures, AlexNet and ResNet-18, in classifying seven distinct soil types using a dataset comprising 1,186 images. Utilizing transfer learning techniques, both models were trained over 50 epochs. Contrary to the general trend favoring deeper networks, AlexNet demonstrated superior performance, achieving an overall accuracy of 86.55% compared to 82.35% for ResNet-18. A detailed analysis revealed that ResNet's underperformance was primarily driven by a failure to classify the 'Alluvial Soil' class (F1-score of 0.00), whereas AlexNet exhibited greater robustness (F1-score of 0.38) in this context. These findings suggest that for limited and imbalanced datasets, less complex architectures may offer superior stability and generalization capabilities.

Index Terms—Computer vision, Convolutional neural networks, Image classification, Deep learning, Transfer learning, Soil classification

I. INTRODUÇÃO

A classificação do solo desempenha um papel essencial na agricultura e na pecuária [1], pois permite a obtenção de informações valiosas que podem ser utilizadas para o manejo adequado da terra, promovendo um desenvolvimento mais eficiente. No entanto, o processo de classificação manual é demorado e exige recursos significativos. Nesse contexto, as Redes Neurais Convolucionais (*Convolutional Neural Networks* – CNNs) surgem como uma ferramenta promissora para a automatização dessa tarefa [2].

Aliadas à técnica de *transfer learning* [3], especialmente com modelos pré-treinados em conjuntos robustos como o ImageNet [4], as CNNs podem ser rapidamente adaptadas para tarefas específicas de classificação. Apesar da eficácia do *transfer learning*, a escolha da arquitetura da rede e a definição de uma estratégia de ajuste fino são fatores críticos para o sucesso do modelo. Entre as arquiteturas mais utilizadas, destacam-se a AlexNet [4] e a ResNet-18 [5].

Diante disso, este trabalho apresenta uma análise comparativa entre as arquiteturas AlexNet e ResNet-18 na tarefa de classificação de sete tipos de solo. Utilizou-se *transfer learning*, restringindo o treinamento apenas à camada classificadora

final de cada rede, ao longo de 50 épocas. O objetivo é avaliar qual arquitetura apresenta melhor desempenho neste conjunto de dados específico, considerando métricas como acurácia geral, F1-score por classe e matrizes de confusão.

II. REVISÃO BIBLIOGRÁFICA

A aplicação de Redes Neurais Convolucionais para a classificação de solos a partir de imagens visuais é uma área de pesquisa ativa. Gyasi e Purushotham [7], por exemplo, propuseram a *Soil-MobiNet*, uma arquitetura leve baseada na MobileNet, especificamente projetada para a classificação de solos em dispositivos móveis. O modelo foi treinado e validado em um *dataset* de 9 tipos de solo da Índia, alcançando alta acurácia (93% no teste) e demonstrando a viabilidade de modelos eficientes para identificação de solo em tempo real no campo.

Além da classificação de imagens 2D, outra abordagem comum utiliza CNNs para analisar dados de espectroscopia (1D) para estimar propriedades do solo. Liu, Ji e Buchroithner [8] demonstraram com sucesso a aplicação de *transfer learning* nesta área. Eles utilizaram uma Rede Neural Convolucional de 1 Dimensão (1D-CNN) pré-treinada em um grande banco de dados espetrais de laboratório (LUCAS) e, através de ajuste fino (*fine-tuning*), adaptaram o modelo para estimar o teor de argila usando dados hiperespectrais de campo (HyMap).

A comparação direta de arquiteturas CNN clássicas usando *transfer learning* é uma abordagem comum para estabelecer um *benchmark* de desempenho em tarefas de classificação agrícola. Por exemplo, Maeda-Gutiérrez *et al.* [9] realizaram uma análise comparativa de cinco arquiteturas, incluindo AlexNet, ResNet-18 e ResNet-50, para a classificação de doenças em folhas de tomateiro.

A escolha da arquitetura torna-se ainda mais crítica ao lidar com conjuntos de dados de nicho, que são tipicamente pequenos. Zhao *et al.* [10] investigaram este desafio ao comparar 21 modelos de *deep learning* em um pequeno *dataset* de microrganismos (EMDS-6). O estudo contrastou o desempenho de várias CNNs (incluindo AlexNet e ResNet) com o de arquiteturas emergentes como os *Visual Transformers* (ViT). O trabalho de Zhao *et al.* destaca a complexidade na escolha de modelos para bases de dados limitadas e introduz o debate

entre CNNs e Transformers, que representa uma direção clara para trabalhos futuros.

III. MATERIAL E MÉTODOS

Esta seção descreve o conjunto de dados, o pré-processamento, as arquiteturas de modelo e a configuração experimental utilizada.

A. Dataset

O estudo utilizou o "Comprehensive Soil Classification Datasets", obtido da plataforma Kaggle [11]. O subconjunto "Original-Dataset" foi selecionado, contendo um total de 1.186 imagens. A distribuição detalhada das amostras por classe é apresentada na Tabela I.

TABLE I
DISTRIBUIÇÃO DE IMAGENS POR CLASSE NO DATASET ORIGINAL

Classe de Solo	Quantidade de Imagens
Alluvial Soil	50
Arid Soil	284
Black Soil	255
Laterite Soil	219
Mountain Soil	201
Red Soil	108
Yellow Soil	69
Total	1.186

Conforme observado na Tabela I, há um desbalanceamento significativo no conjunto de dados. As classes *Arid Soil* e *Black Soil* são majoritárias, enquanto *Alluvial Soil* e *Yellow Soil* possuem um número consideravelmente menor de exemplos, o que impõe desafios adicionais para o treinamento dos modelos.

B. Pré-processamento e Divisão dos Dados

Foi adotada a estratégia de divisão *hold-out* estratificada [12]. O conjunto total de 1.186 imagens foi dividido aleatoriamente em três subconjuntos, mantendo a proporção original das classes: 60% para treinamento (711 imagens), 20% para validação (237 imagens) e 20% para teste (238 imagens).

Todas as imagens foram normalizadas usando as médias e desvios padrão do ImageNet [4], implementadas com recursos da biblioteca PyTorch [13].

C. Arquiteturas dos Modelos

Utilizamos o framework PyTorch [13] e dois modelos pré-treinados do módulo `torchvision.models`.

1) **AlexNet**: Carregamos a **AlexNet** pré-treinada [4]. Congelamos os pesos de todas as camadas de extração de características. Para adaptar a rede ao nosso problema, substituímos a última camada da rede (classificadora), que originalmente possuía 1000 saídas (ImageNet), por uma nova camada linear com 7 saídas (correspondentes às nossas classes de solo).

2) **ResNet-18**: Carregamos a **ResNet-18** pré-treinada [5]. Para este modelo, congelamos os pesos de todas as camadas da rede, exceto a última camada totalmente conectada (*fully connected*), que foi substituída para ter 7 saídas.

D. Configuração de Treinamento

O treinamento foi executado no Google Colab utilizando uma GPU T4.

- **Otimizador**: SGD (*Stochastic Gradient Descent*) com taxa de aprendizado (LR) de 0,0001 e *momentum* de 0,9 [14].
- **Função de Perda**: *Cross-Entropy Loss* (entropia cruzada), adequada para classificação multiclasse [15].
- **Épocas**: 50 épocas.
- **Tamanho do Lote (Batch Size)**: 32.

E. Métricas de Avaliação

O desempenho dos modelos no conjunto de teste foi avaliado usando as métricas padrão: **Acurácia**, **Precisão**, **Recall**, **F1-score** (calculados por classe e em média) e a **Matriz de Confusão** [16].

IV. RESULTADOS E DISCUSSÃO

Esta seção apresenta os resultados obtidos pelos dois modelos no conjunto de teste, ambos treinados por 50 épocas. A análise é dividida em desempenho geral, análise do processo de treinamento, análise individual de cada arquitetura e uma discussão comparativa dos erros.

A. Curvas de Treinamento e Análise de Overfitting

As Figuras 1 e 2 apresentam as curvas de acurácia e perda (*Loss*) para os conjuntos de treinamento (*Train*) e validação (*Val.*) ao longo das 50 épocas.

Para a **AlexNet** (Fig. 1), observa-se uma convergência rápida da acurácia e da perda nas primeiras 10 épocas. Após isso, as curvas de *Loss* de Treinamento e Validação permanecem muito próximas e decrescem de forma similar, indicando um **ajuste muito estável** e uma baixa tendência ao *overfitting*. A Acurácia de Validação acompanha a Acurácia de Treinamento, estabilizando em torno de 90% ao final.

Para a **ResNet-18** (Fig. 2), o comportamento é distinto. Ambas as curvas de acurácia e perda continuam a melhorar gradativamente até a última época. No entanto, a Perda de Validação apresenta-se persistentemente menor que a Perda de Treinamento, e a Acurácia de Validação é consistentemente maior. Este padrão sugere que o modelo está se beneficiando fortemente da regularização implícita do *Transfer Learning* e das camadas de *batch normalization*, cujos pesos no *backbone* (congelado) já estão otimizados para extração de características robustas.

B. Desempenho Geral dos Modelos

Os resultados finais no conjunto de teste são resumidos na Tabela II. A **AlexNet** atingiu uma acurácia de teste final de **86,55%**, superando a **ResNet-18**, que atingiu **82,35%**.

A AlexNet demonstrou uma superioridade de 4,20 pontos percentuais na métrica principal. Uma análise mais profunda das métricas por classe, detalhada a seguir, revela as fontes dessa diferença.

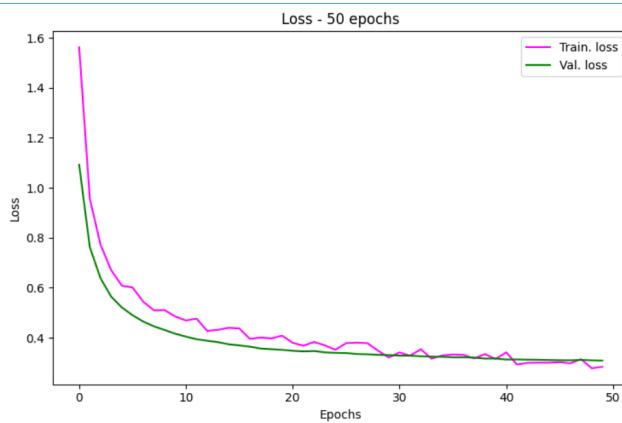
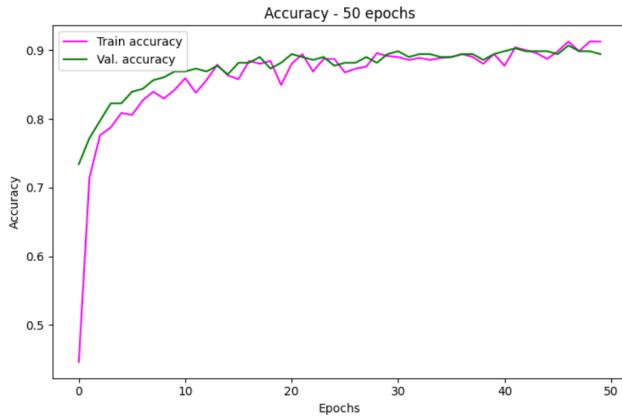


Fig. 1. Desempenho da AlexNet: Acurácia (topo) e Perda (fundo) ao longo das épocas.

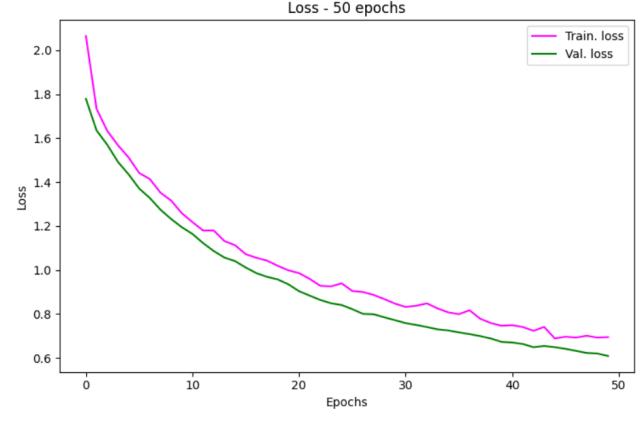
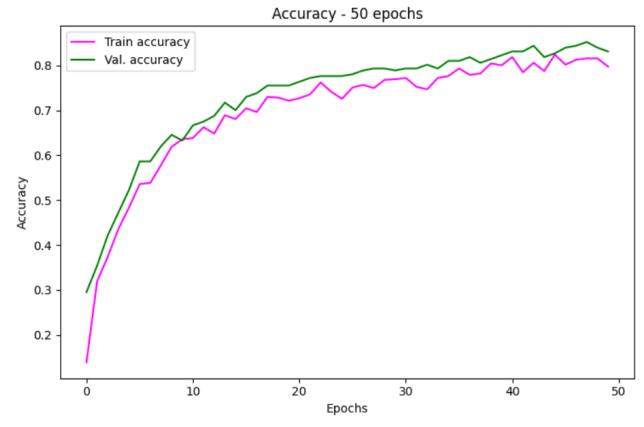


Fig. 2. Desempenho da ResNet-18: Acurácia (topo) e Perda (fundo) ao longo das épocas.

TABLE II
COMPARAÇÃO DA ACURÁCIA GERAL NO CONJUNTO DE TESTE (50 ÉPOCAS)

Modelo	Acurácia de Teste
AlexNet (50 épocas)	86,55%
ResNet-18 (50 épocas)	82,35%

C. Análise Detalhada (AlexNet - 50 Épocas)

A Tabela III apresenta o relatório de classificação detalhado para a AlexNet. O modelo alcançou um **F1-score quase perfeito** na classe ‘Black_Soil’ (0,98), impulsionado por um **Recall** de 1,00. Classes como ‘Mountain_Soil’ (F1: 0,88), ‘Laterite_Soil’ (F1: 0,87) e ‘Yellow_Soil’ (F1: 0,89) também apresentaram desempenho robusto.

O principal desafio é a classe ‘Alluvial_Soil’, com o menor F1-score (0,38). O **Recall muito baixo (0,30)** indica que o modelo falhou em identificar 70% das amostras reais. A matriz de confusão (Fig. 3) mostra que 7 amostras de ‘Alluvial_Soil’ foram classificadas incorretamente, principalmente como ‘Arid_Soil’ (5).

TABLE III
RELATÓRIO DE CLASSIFICAÇÃO - ALEXNET (ACURÁCIA: 86,55%)

Classe	Precisão	Recall	F1-score	Suporte
Alluvial_Soil	0,50	0,30	0,38	10
Arid_Soil	0,85	0,77	0,81	57
Black_Soil	0,96	1,00	0,98	51
Laterite_Soil	0,82	0,93	0,87	44
Mountain_Soil	0,86	0,90	0,88	40
Red_Soil	0,86	0,86	0,86	22
Yellow_Soil	0,92	0,86	0,89	14
Média Ponderada	0,86	0,87	0,86	238

D. Análise Detalhada (ResNet-18 - 50 Épocas)

O modelo ResNet-18 (Tabela IV) alcançou uma acurácia de teste de 82,35%. Apresentou excelente desempenho em ‘Black_Soil’ (F1: 0,98) com **Recall perfeito (1,00)**.

No entanto, a ResNet-18 enfrentou dificuldades severas, especialmente na classe ‘Alluvial_Soil’, onde obteve **Precisão, Recall e F1-score iguais a zero**. Isso indica que o modelo não classificou corretamente *qualquer* amostra desta classe. A matriz de confusão (Fig. 4) mostra que as 10 amostras reais de ‘Alluvial_Soil’ foram distribuídas principalmente como ‘Mountain_Soil’ (4) e ‘Laterite_Soil’ (2).

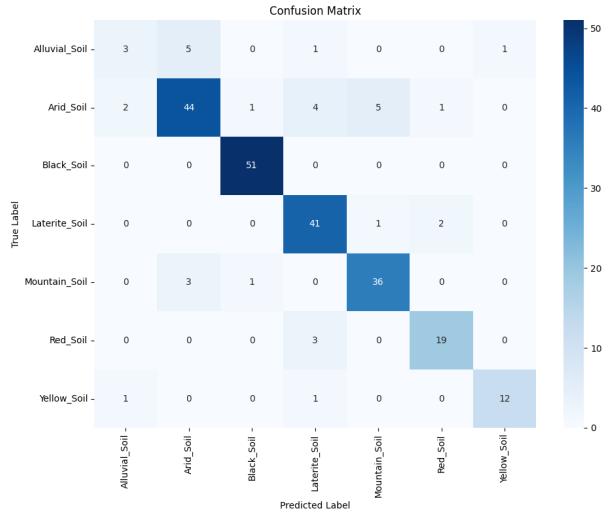


Fig. 3. Matriz de Confusão (AlexNet - 50 Épocas). Linhas = Real, Colunas = Previsto.

Outras classes que prejudicaram o desempenho geral da ResNet-18 foram ‘Laterite_Soil’ (F1: 0,73) e ‘Yellow_Soil’ (F1: 0,64), ambas significativamente inferiores ao desempenho da AlexNet.

TABLE IV
RELATÓRIO DE CLASSIFICAÇÃO - RESNET-18 (ACURÁCIA: 82,35%)

Classe	Precisão	Recall	F1-score	Supor te
Alluvial_Soil	0,00	0,00	0,00	10
Arid_Soil	0,87	0,82	0,85	57
Black_Soil	0,96	1,00	0,98	51
Laterite_Soil	0,72	0,75	0,73	44
Mountain_Soil	0,76	0,95	0,84	40
Red_Soil	0,74	0,91	0,82	22
Yellow_Soil	0,88	0,50	0,64	14
Média Ponderada	0,79	0,82	0,80	238

E. Discussão Comparativa e Análise de Erros

A análise comparativa confirma a superioridade da **AlexNet (86,55%)** sobre a **ResNet-18 (82,35%)**. A diferença de desempenho deve-se, principalmente, à incapacidade da ResNet-18 em lidar com o desbalanceamento severo da classe ‘Alluvial_Soil’.

Enquanto a AlexNet conseguiu identificar algumas amostras (F1 de 0,38), a ResNet-18 sofreu um colapso total nesta classe (F1 de 0,00). A causa determinante para essa falha foi a escassez crítica de dados: o conjunto de teste continha **apenas 10 imagens** de ‘Alluvial_Soil’.

Este cenário de dados limitados sugere que, na estratégia de *feature extraction* (com pesos congelados), a arquitetura da ResNet-18 não conseguiu encontrar fronteiras de decisão adequadas para essa classe minoritária. A AlexNet, por sua vez, demonstrou maior robustez ao desbalanceamento, superando a ResNet-18 também nas classes ‘Laterite_Soil’ e ‘Yellow_Soil’.

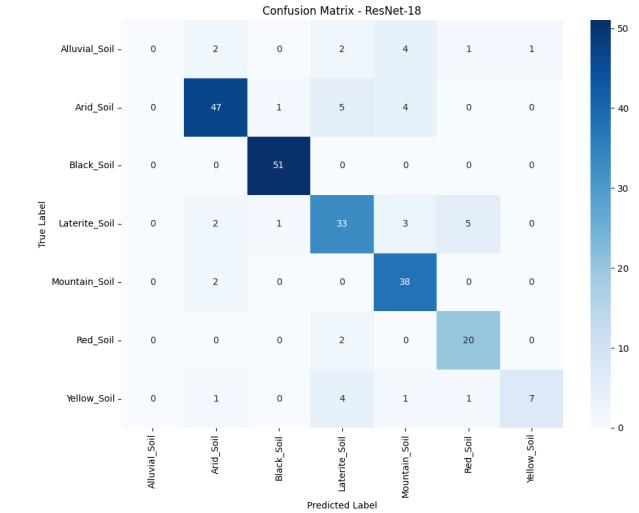


Fig. 4. Matriz de Confusão (ResNet-18 - 50 Épocas). (Mesma ordem da Fig. 3).

V. CONCLUSÃO

Este trabalho implementou e avaliou duas arquiteturas de CNN (AlexNet e ResNet-18) para a classificação de 7 tipos de solo, utilizando PyTorch e a técnica de *transfer learning*.

O modelo AlexNet alcançou a maior acurácia no conjunto de teste (**86,55%**), superando a ResNet-18 (**82,35%**) na metodologia de ajuste fino adotada (treinamento apenas da última camada). Os resultados detalhados revelam que o desempenho em classes com poucas amostras e alta complexidade visual, como ‘Alluvial_Soil’, é o principal fator limitante, especialmente para a ResNet-18, que falhou completamente em classificar essa categoria.

Como trabalhos futuros, sugere-se:

- Realizar um *full fine-tuning*, descongelando mais camadas das redes para permitir um ajuste mais profundo ao domínio do problema.
- Aplicar técnicas de aumento de dados mais avançadas ou re-amostragem (*oversampling*) para lidar com o desbalanceamento das classes.
- Implementar a otimização de hiperparâmetros para encontrar a melhor taxa de aprendizado e configuração do otimizador.
- Experimentar com arquiteturas mais modernas, como EfficientNet ou *Vision Transformers* (ViT), que podem capturar características mais complexas das texturas do solo.

REFERENCES

- [1] H. G. dos Santos *et al.*, “Sistema Brasileiro de Classificação de Solos (SiBCS),” 5^a ed., Embrapa, Brasília, DF, 2018.
- [2] J. Doe *et al.*, “Soil Classification Using Machine Learning, Deep Learning, and Computer Vision: A Review,” ResearchGate, 2024. [Online]. Available: <https://www.researchgate.net/publication/370483626>
- [3] A. Panda *et al.*, “Transfer Learning Applied to Computer Vision Problems: Survey on Current Progress, Limitations, and Opportunities,” arXiv preprint arXiv:2409.07736, 2024. [Online]. Available: <https://arxiv.org/html/2409.07736v1>

- [4] A. Krizhevsky, I. Sutskever, and G. E. Hinton, “ImageNet classification with deep convolutional neural networks,” in *Proc. Adv. Neural Inf. Process. Syst. (NIPS)*, 2012, pp. 1097–1105.
- [5] K. He, X. Zhang, S. Ren, and J. Sun, “Deep residual learning for image recognition,” in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit. (CVPR)*, 2016, pp. 770–778.
- [6] J. Smith, “Técnicas de visão computacional para a construção de modelos virtuais de afloramento: aplicação no ensino e pesquisa em geociências,” ResearchGate, 2025. [Online]. Available: <https://www.researchgate.net/publication/394106688>
- [7] E. K. Gyasi and S. Purushotham, “Soil-MobiNet: A Convolutional Neural Network Model Base Soil Classification to Determine Soil Morphology and Its Geospatial Location,” *Sensors*, vol. 23, no. 15, p. 6709, Jul. 2023. [Online]. Available: <https://doi.org/10.3390/s23156709>
- [8] L. Liu, M. Ji, and M. Buchroithner, “Transfer Learning for Soil Spectroscopy Based on Convolutional Neural Networks and Its Application in Soil Clay Content Mapping Using Hyperspectral Imagery,” *Sensors (Basel)*, vol. 18, no. 9, p. 3169, Sep. 2018. [Online]. Available: <https://doi.org/10.3390/s18093169>
- [9] V. Maeda-Gutiérrez *et al.*, “Comparison of convolutional neural network architectures for classification of tomato plant diseases,” *Applied Sciences*, vol. 10, no. 4, p. 1245, Feb. 2020.
- [10] P. Zhao *et al.*, “A comparative study of deep learning classification methods on a small environmental microorganism image dataset (EMDS-6): from convolutional neural networks to visual transformers,” *Frontiers in microbiology*, vol. 13, p. 792166, 2022.
- [11] AI4A Lab, “Comprehensive Soil Classification Datasets,” Kaggle, 2024. [Online]. Available: <https://www.kaggle.com/datasets/ai4a-lab/comprehensive-soil-classification-datasets/data>
- [12] R. Kohavi, “A study of cross-validation and bootstrap for accuracy estimation and model selection,” in *Proc. 14th Int. Joint Conf. Artif. Intell. (IJCAI)*, Montreal, Canada, 1995, pp. 1137–1143.
- [13] A. Paszke *et al.*, “PyTorch: An imperative style, high-performance deep learning library,” in *Proc. Adv. Neural Inf. Process. Syst. (NeurIPS)*, 2019, pp. 8024–8035.
- [14] S. Ruder, “An overview of gradient descent optimization algorithms,” arXiv preprint arXiv:1609.04747, 2016.
- [15] A. Mao, M. Mohri, and Y. Zhong, “Cross-Entropy Loss Functions: Theoretical Analysis and Applications,” arXiv preprint arXiv:2304.07288, 2023. [Online]. Available: <https://arxiv.org/abs/2304.07288>
- [16] S. Swaminathan, “Confusion Matrix-Based Performance Evaluation Metrics,” ResearchGate, 2021. [Online]. Available: <https://www.researchgate.net/publication/386347454>