

A Comparison Study of Deep Convolutional Neural Networks for the Classification of Brazilian Savannah Pollen Grains: Preliminary Results

Bruno Aristimunha¹, Felipe Silveira Brito Borges^{2*}, Ariadne Barbosa Gonçalves², Hemerson Pistori²

¹Universidade Federal do ABC, UFABC, Santo André, SP, Brasil

²Universidade Católica Dom Bosco, UCDB, Campo

b.aristimunha@gmail.com, *eng.fe.silveira@gmail.com, ariadne.gon@gmail.com, pistori@ucdb.br

Resumo—The classification of pollen grains images are currently done manually and visually, being a weariful task and predisposed to mistakes due to human exhaustion. In this paper, the authors introduce an automatic classification of 55 different pollen grain classes, using convolutional neural networks. Different architectures and hyperparameters were used to improve the classification result. Using the networks VGG16, VGG19, and InceptionV3, were obtained accuracy rates over 93.58%.

Index Terms—Computer Vision, Machine Learning, Palynology, Pollen Grains.

I. INTRODUÇÃO

A palinologia é a ciência que estuda os grãos de pólen e esporos, estes são objetos microscópicos produzidos pelas plantas classificadas como: briófitas, pteridófitas, gimnospermas e angiospermas, ou seja, são células reprodutivas provenientes das flores possuindo um tamanho médio de 10 a 50 μ metros [Joosten and de Klerk, 2002]. Ademais, outras áreas de conhecimento ramificaram-se da palinologia, como o estudo de polens de mel através da melissopalynologia [Von Der Ohe et al., 2004], o estudo de polens causadores de alergia na imunologia [Delves et al., 2017] e o estudo de ecossistemas do passado através da paleoecologia [Jackson and Williams, 2004].

Devido a dimensão dos grãos de pólen, para a sua visualização é necessário o uso de um microscópio para a captura de imagens e posterior análise. Existem diversas maneiras de analisar os grãos de pólen, como a identificação através de microscópio [Collinvaux et al., 2014], a análise por varredura eletrônica [Nakagawa et al., 2000] e, ainda, a identificação por partículas a laser [Boyain-Goitia et al., 2003].

A identificação destes grãos por olho humano, através de um microscópio é o método mais comum, pois não há necessidade de um grande investimento financeiro, pois utilizam-se apenas lâminas palinológicas contendo as amostras de grãos, além de um especialista da área. Em contrapartida, muitos erros são comumente cometidos neste tipo de análise, sendo causados pelo desgaste exercido na realização da tarefa de classificação. Outro ponto importante é a baixa demanda de especialistas da área da palinologia, criando uma dependência devido a falta de profissionais qualificados para tal tarefa.

A atividade de classificar e contar grãos de pólen pode ser automatizada com o uso da visão computacional e da aprendizagem de máquina [Hwang et al., 2013]. De acordo com [Prince, 2012], o objetivo da visão computacional é desenvolver computacionalmente a habilidade visual que os seres humanos já possuem e tornar a tarefa mais rápida e eficiente, utilizando-se de algoritmos e técnicas computacionais desenvolvidas para determinado objetivo. Essa área tem como foco automatizar tarefas associadas a visão humana. No caso deste artigo, visa automatizar a classificação feita visualmente por especialistas da área da palinologia, corretamente e de forma mais rápida e eficaz [Quinta, 2013].

Para automatizar o trabalho exercido com o uso da visão computacional, destinado a identificação das imagens de grãos, é feita a combinação com a inteligência artificial. O aprendizado profundo é um ramo da inteligência artificial destinado para problemas de percepção, como neste caso, a identificação de imagens. Segundo [Ng, 2019], este tipo de aprendizado se destaca quando o assunto é uma grande quantidade de dados e problemas complexos a serem resolvidos, sendo uma família de técnicas que se adaptam a estes. Analogamente, o seu uso para identificar grandes quantidades de imagens, onde estão presentes diferentes famílias e espécies de grãos de pólen, torna-se adequado.

No presente trabalho, os autores desenvolveram um classificador automático de polens utilizando diferentes Redes Neurais do Aprendizado Profundo, que são amplamente utilizadas na literatura, constituindo-se o estado da arte. As redes utilizadas foram: VGG16, VGG19, InceptionV3 e ResNet50, com o objetivo de classificar corretamente 55 diferentes classes de grãos de pólen.

Este artigo está organizado da seguinte forma: A Seção II apresenta os trabalhos correlatos. Os materiais e métodos adotados nesta pesquisa estão descritos na Seção III. A Seção IV relata os resultados obtidos e a discussão. Por fim, as conclusões e trabalhos futuros são mostrados na Seção V.

II. TRABALHOS CORRELATOS

De acordo com [Quinta et al., 2012], o reconhecimento de padrões pode ser modelado com o uso de Floresta de Cami-

nhos Ótimos. Experimentos foram executados com diferentes conjuntos de imagens de grãos de pólen, sendo um conjunto com 4 classes e outros três com 6 classes. O melhor resultado utilizando essa técnica foi 75.27% de acurácia utilizando o algoritmo C4.5.

O uso da técnica Bag of Words [Gonçalves et al., 2013] obteve o melhor desempenho através do algoritmo SMO, que resultou em 71% de acerto na métrica de percentual de classificação correta. Esse experimento utilizou 6 classes de grãos de pólen. Sob o mesmo ponto de vista, a aplicação da Transformada Wavelet [Da Silva et al., 2014], na classificação de grãos de pólen com os algoritmos C4.5 e o KNN, obteve resultados não muitos bons, sendo 57% de taxa na medida-F com o algoritmo C4.5, porém, quando o classificador foi combinado com mais atributos extraídos, o resultado de classificação obteve uma melhora significativa, sendo 79% de taxa na medida-F. Foram utilizadas 7 espécies de grãos de pólen neste trabalho.

Em contrapartida, de acordo com [Gonçalves et al., 2016], a melhor técnica que deve ser usada para automatizar a classificação de grãos de polens é a combinação dos atributos de cor, forma e textura com o Bag of Words, além de, com o classificador CSVC executar a classificação. A partir dos algoritmos de aprendizagem supervisionada e com um banco de imagens com 9 classes de grãos de pólen, o melhor resultado obtido foi através do experimento CST+ BOW com o algoritmo SVM, obtendo 79% na métrica de percentual de classificação correta.

Ainda segundo [Gonçalves et al., 2016], a classificação visual de grãos de pólen, através de especialistas da área atinge a margem de 64% de classificação correta (CCR). Já algoritmos de aprendizagem de máquina, como, por exemplo, o CST+BOW, atingem o valor de 97%. Neste experimento, foi utilizado um banco de imagens de 23 espécies de grãos de pólen.

III. MATERIAIS E MÉTODOS

As próximas subseções cobrem os materiais e métodos empregados neste artigo.

A. Banco de Imagens

Foi utilizado um banco de imagens de polens nativos do Mato Grosso do Sul. Por se tratar de um objeto de geometria tridimensional, as diferentes angulações influenciam tanto na classificação do especialista quanto em algoritmos tradicionais de classificação, portanto cada grão possuía 35 imagens distintas.

O conjunto de dados possui 55 classes balanceadas, com um total de 1925 imagens. Técnicas de aumento de dados nas imagens foram largamente empregadas. Optou-se pelo aumento de dados empregando rotações de 45°, traçando um paralelo com a literatura botânica, que emprega rotações de 45° na captura de imagens. Por se tratar de um objeto de geometria tridimensional levantou-se a hipótese de que essas rotações auxiliariam na generalização dos métodos.

Foi considerada a divisão: treino, validação e teste, sendo 60, 20 e 20 a porcentagem de cada uma de acordo com a quantidade de imagens. Na Figura 1 são ilustradas as 55 classes utilizadas neste trabalho. Todas as figuras foram redimensionadas para 299 por 299 pixels.

B. Configurações

Todos os modelos foram implementados a partir do framework Keras¹, com back-end do TensorFlow. Os experimentos foram realizados usando uma CPU com Intel Core i7-5930K de 3.50 GHz e duas GPUs: Nvidia Quadro K5200 e uma GeForce GTX 970.

Inicialmente, foram empregadas três arquiteturas distintas: InceptionV3, VGG16 e VGG19. Nesses modelos, a otimização do gradiente descendente estocástico com taxa de aprendizado de 10^{-3} e impulso de 0.9, foram utilizados, além de serem empregadas 100 épocas de treinamento.

Os experimentos foram executados com a técnica de Ajuste Fino - AF, em que os pesos dos modelos foram inicializados com a rede já previamente treinados no banco de imagens ImageNet. Para cada arquitetura, houve o re-treino parcial de diferentes porcentagens das camadas: 0%, 25%, 50%, 75% e 100%. Portanto, em 0% foram utilizados todos os pesos do ImageNet sem nenhum re-treino com as imagens de pólen; em 25% houve o re-treino das camadas no último percentil, logo, as 75% das camadas no início não sofreram atualização nos pesos; para 50% os pesos de apenas segunda metade das camadas foram atualizados; e 100% houve o re-treino dos pesos utilizando o banco de pólen, logo Transferência de Aprendizado - TA. Houve também o re-treino total sem a inicialização de pesos, ou seja, sem Transferência de Aprendizado - Sem TA.

Além disso, dado o tamanho amostral do conjunto de dados, optou-se também na verificação da melhor taxa de aprendizado. Para tanto, a rede ResNet50, com re-treino de 25% das camadas finais com inicialização pesos previamente treinado na ImageNet, foi utilizada. Somente nessa arquitetura foi analisada a influência de diferentes taxas de aprendizado e uma regra de otimização adequada para conjuntos pequenos de Adam [Kingma and Ba, 2014].

Para o método de otimização gradiente descendente estocástico, a perda e acurácia foram os seguintes valores: $\{10^{-2}, 10^{-3}, 10^{-4}$ e $10^{-5}\}$. Já para o regra de otimização Adam, os valores nativos $\alpha = 0.001$, $\epsilon = 10^{-8}$, $\beta_1 = 0.9$ e $\beta_2 = 0.999$ foram utilizados.

As seguintes camadas foram adicionadas nas arquiteturas: achatamento, totalmente conectada, *dropout*, totalmente conectada, *dropout* e totalmente conectada para predição com ativação *Softmax*.

C. Métricas

Para todas as arquiteturas, utilizou-se na função de custo a entropia cruzada, conforme a Equação 1:

¹Versão: 2.1.1

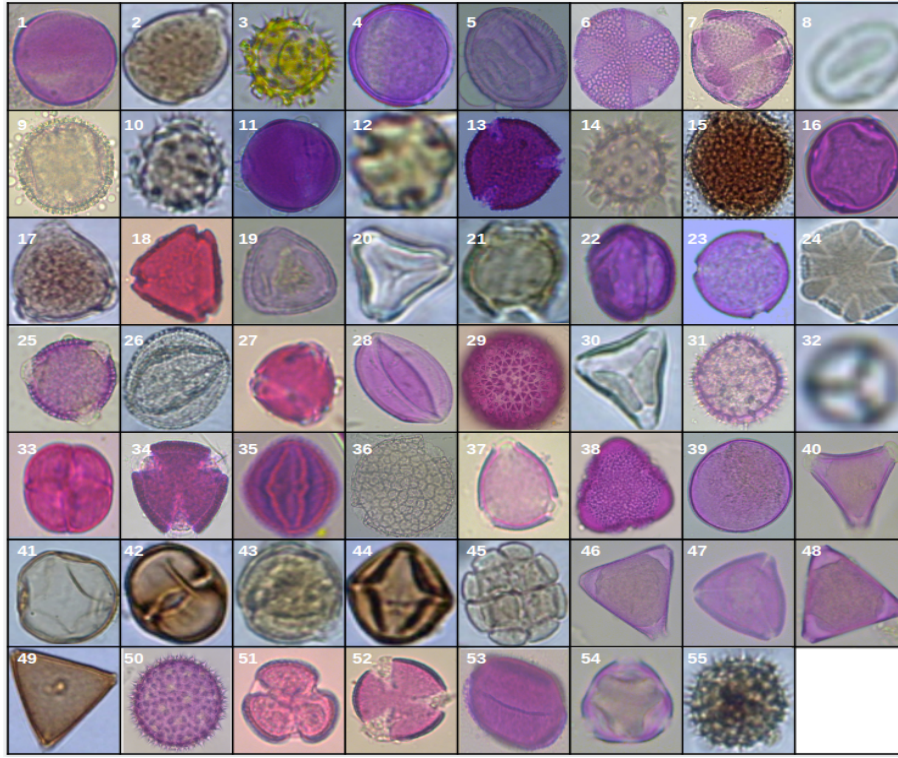


Figura 1. Exemplo de imagens para cada um das 55 espécies de pólen utilizadas. (Escala = 10 μ metros). 1-*Arachis sp.*, 2-*Arrabidaea florida*, 3-*Aspilia gracielae*, 4-*Bacopa australis*, 5-*Beladona*, 6-*Caesalpinia peltophoroïdes*, 7-*Caryocar brasiliensis*, 8-*Cecropia pachystachya*, 9-*Ceiba speciosa*, 10-*Chromolaena laevigata*, 11-*Cissus campestris*, 12-*Combretum discolor*, 13-*Cordia trichotoma*, 14-*Cosmos caudatus*, 15-*Croton urucurana*, 16-*Curva lago*, 17-*Dipteryx alata*, 18-*Doliodocarpus dentatus*, 19-*Erythina dominguesi*, 20-*Eucalyptus sp.*, 21-*Fareamea sp.*, 22-*Genipa aumiricana*, 23-*Guazuma ulmifolia*, 24-*Hypitis sp.*, 25-*Ligustrum lucidum*, 26-*Mabea fistulifera*, 26-*Machaerium aculeatum*, 28-*Magnolia champaca*, 29-*Manihot esculenta*, 30-*Matayba guianensis*, 31-*Mauritia flexuosa*, 32-*Mimosa ditans*, 33-*Mimosa pigra*, 34-*Mitostemma brevifilis*, 35-*Myracrodunon urundeuva*, 36-*Ochroma*, 37-*Ouratea exasperma*, 38-*Pachia aquatica*, 39-*Palmeira real*, 40-*Paullinia spicata*, 41-*Poaceae*, 42-*Protium heptaphyllum*, 43-*Qualea multiflora*, 44-*Schinus sp.*, 45-*Senegalia plumosa*, 46-*Serjania erecta*, 47-*Serjania hebecarpa*, 48-*Serjania laruotteana*, 49-*Serjania sp.*, 50-*Sida cerradoensis*, 51-*Tabebuia chysotricha*, 52-*Tabebuia rosealba*, 53-*Tradescantia Pallida*, 54-*Trembleya phlogiformis* e 55-*Tridax procumbens*.

$$loss(p, q) = -\frac{1}{|I|} \sum_{i \in I} \sum_{x \in C} p_i(x) \log(q_i(x)) \quad (1)$$

em que $q_i(x)$ é a probabilidade estimada da saída pertencer a classe x , e $p_i(x)$ é a real probabilidade. Considerando o balanceamento das classes e convergência da função, ao utilizar essa função, obtém-se suficiente informação, pois há uma maximização de acertos na classificação dos grãos de pólen em relação a outros durante o treino. Foi também verificado a precisão do método durante o treinamento. Além disso, nos testes foi analisada a medida F_1 , que é descrita como:

$$F_1 = \frac{2 \cdot Precisão \cdot Revocacao}{Precisão + Revocacao} \quad (2)$$

Ela consiste em uma média harmônica entre a precisão e a revocação, em outras palavras, uma forma de verificar em um único valor a qualidade geral dos modelos. Assim sendo, um modelo perfeito possui:

$$F_1 = \frac{2 \cdot 1 \cdot 1}{1 + 1} = 1. \quad (3)$$

D. Descrição das arquiteturas

Nessa subseção, são brevemente descritos os funcionamentos de cada arquitetura empregada nesse artigo.

1) *Visual Geometry Group - VGG*: A Visual Geometry Group projeta uma rede simples para reconhecimento de imagem. Há apenas o agrupamento de 3 por 3 e o agrupamento de 2 por 2. Os números 16 e 19 dizem respeito ao número de camada. Essa arquitetura consiste em uma pilha de camadas convolucionais, basicamente sendo um conjunto de 3 camadas convolucionais de 3 por 3 pixels, com passo de 1 pixel, depois um preenchimento de 1 pixel, dessa forma se preserva a resolução espacial após convolução.

Subseqüentemente realiza-se um Maxpool através de uma janela 2 por 2, com stride 2, seguindo de 5 camadas iguais à essa temos uma dimensão 7x7x512. Após essas camadas há uma camada totalmente conectada, e por último um classificador SoftMax.

2) *InceptionV3*: Conhecida como GoogLeNet, posteriormente foi rebatizada como Inception vN, com $N \in 1, 2, 3, 4$ correspondente à sua versão. A rede Inception diverge da rede VGG pela sua alta especialização das camadas.

O GoogLeNet contém 9 módulos de inicialização empilha-

dos linearmente, 27 camadas de profundidade considerando camadas de pooling.

A principal premissa dessa arquitetura é que existe variação na posição e tamanho dos objetos dentro da imagem. Assim sendo, deve-se integrar à rede de convoluções tamanhos de filtros diferentes, uma vez que, computacionalmente, não é possível calcular tamanhos distintos. Escolhe-se ainda, delimitar tamanhos de filtro dentro das camadas ocultas: 1X1, 3x3, 5x5.

Ao realizar esses múltiplos tamanhos de filtro, garante-se que, caso em algum dos recortes haja uma figura, durante o aprendizado haverá o retorno positivo para aquele local.

A versão 3 da arquitetura consiste de atualizações para simplificar a complexidade computacional, melhora na precisão numérica dentro da arquitetura, convoluções fatoradas de 7x7 e normalização de lote nos classificadores auxiliares.

3) *ResNet*: Essa arquitetura foi a primeira a ultrapassar o erro humano na classificação do conjunto de imagem do ImageNet [Russakovsky et al., 2015]. Possui conexões saltadas especiais e um intensivo uso de normalização do lote (*BatchNormalization*).

Essa arquitetura, assim como a InceptionNet não possui camadas totalmente conectadas no final da rede. A motivação de surgimento das ResNet vem de uma problemática no empilhamento de uma rede profunda. Redes como VGG e Inception não melhoram na medida do empilhamento e, espera-se que na medida da profundidade, as redes convulsionais adquirissem mais performance. De forma prática, uma rede profunda como VGG com 28 camadas apresenta um resultado melhor do que uma com 56 camadas. A generalização do modelo não aumenta com o incremento de profundidade. Situações como essas podem aparentar *overfitting*, no entanto, não apresentam comportamento divergentes no teste e treino. A hipótese geradora do método ResNet caracteriza isso como um problema de otimização.

Modelos muito profundos são difíceis de otimizar considerando suas múltiplas camadas e conexões locais. A solução encontrada consiste em copiar a camada aprendida no início e configurar as camadas adicionais de identificação. Além disso, nessa arquitetura, uma camada de convolução no começo e ao final da camada totalmente conectada foi adicionada. Por preservar o tamanho das camadas ocultas, torna-se possível generalizar a profundidade facilmente para 34, 50, 101 ou 152 camadas. Outro ponto importante dessa rede em comparação com as demais é a remoção de camadas de *dropout*.

IV. RESULTADOS E DISCUSSÃO

Nesta seção são descritos os resultados obtidos pela metodologia proposta. Para a avaliação das arquiteturas, são descritos os valores obtidos na função de perda, a acurácia e valores da Medida-F.

A. Avaliação das Arquiteturas

Na Tabela I, são observados os valores finais da função de perda, e na Tabela II, encontra-se a precisão final do método no conjunto de treino e validação.

Retreino Arquitetura	AF 0%	AF 25%	AF 50%	AF 75%	TA 100%	Sem TA
InceptionV3	0.25 0.19	0.27 0.26	0.34 0.27	0.56 0.45	1.09 0.76	1.87 1.49
VGG16	0.39 0.37	0.34 0.35	0.49 0.46	0.83 0.72	2.27 2.00	4.01 4.01
VGG19	0.35 0.37	0.46 0.49	0.51 0.60	0.83 0.77	2.67 2.41	4.01 4.01

Tabela I

VALORES DA FUNÇÃO DE PERDA NA ÚLTIMA ÉPOCA DO TREINAMENTO E DA VALIDAÇÃO. POR REDE TEMOS NA LINHA SUPERIOR O VALOR NO TREINAMENTO, E NA LINHA INFERIOR NA VALIDAÇÃO. AS CONFIGURAÇÕES FORAM AJUSTE FINO - AF EM 0%, 25%, 50%, 75% E 100% (TRANSFERÊNCIA DE APRENDIZADO - TA), E SEM TRANSFERÊNCIA DE APRENDIZADO - SEM TA.

Retreino Arquitetura	AF 0%	AF 25%	AF 50%	AF 75%	TA 100%	Sem TA
InceptionV3	92.24 93.54	91.90 92.02	88.70 91.83	81.48 84.98	67.46 77.38	36.69 51.52
VGG16	86.18 86.88	89.04 88.22	83.53 85.17	72.36 75.10	35.06 50.38	4.15 6.08
VGG19	87.75 87.26	84.31 88.41	83.12 80.99	73.18 76.24	25.12 38.21	3.27 4.75

Tabela II

VALORES DA ACURÁCIA NA ÚLTIMA ÉPOCA DO TREINAMENTO E DA VALIDAÇÃO. POR REDE TEMOS NA LINHA SUPERIOR O VALOR NO TREINAMENTO, E NA LINHA INFERIOR NA VALIDAÇÃO. AS CONFIGURAÇÕES FORAM AJUSTE FINO - AF EM 0%, 25%, 50%, 75% E 100% (TRANSFERÊNCIA DE APRENDIZADO - TA), E SEM TRANSFERÊNCIA DE APRENDIZADO - SEM TA.

Em relação a função de perda e precisão associadas a cada arquitetura, observou-se que os menores valores se encontram nas primeiras configurações avaliadas, em outras palavras, quanto menor a porcentagem de ajuste fino. Isso ocorre em decorrência do tamanho do conjunto de dados.

Na Tabela III há os valores da medida-F no conjunto de teste. A Figura 2 expõem o resultado das arquiteturas para a métrica medida F.

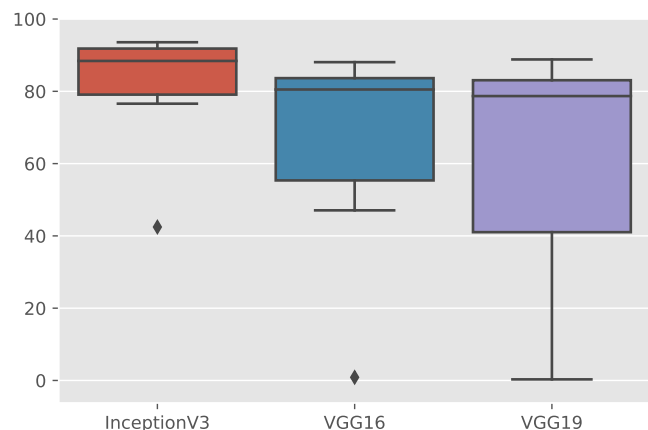


Figura 2. Diagrama de caixa comparando desempenho entre as arquiteturas para a Medida-F.

Segundo a Figura 2, a arquitetura VGG19 apresenta a maior variação e mediana alta. Desse modo, apesar da arquitetura apresentar uma mediana alta, os valores não são tão consis-

Rede	Medida F					
	AF0	AF25	AF50	AF75	TA	S-TA
InceptionV3	93.58	92.37	90.20	86.61	76.58	42.46
VGG16	84.59	88.07	80.82	80.19	47.07	0.87
VGG19	88.82	82.80	83.16	74.57	29.85	0.30

Tabela III
VALORES PARA MEDIDA F NO CONJUNTO DE TESTE.

tentes quanto os valores do VGG16.

A partir da análise da relação entre métricas no treino-validação-teste, observadas nas Tabela anteriores, averiguou-se que há uma forte correlação entre elas. A precisão, a perda e a F_1 possuem um padrão de melhores resultados. A partir do cálculo da correlação de Person entre elas, foram observados os seguintes valores, para F_1 x Precisão no Treino: **98.98%**, F_1 x Perda no Treino: **99.25%**; Para F_1 x Precisão na Validação: **99.55%**; Para F_1 x Perda na Validação: **98.71%**.

A arquitetura InceptionV3 se sobressaiu das demais, pela medida-F com uma baixa variância nos resultados, dentro dos hiper-parâmetros buscados. Ou seja, as informações buscadas não estão em uma posição uniforme da imagem. Isto é condizente com as dificuldades experimentais encontradas para construção do banco de imagens, tendo em vista os objetos analisados se encontram em diferentes solventes e que compete ao especialista da área enquadrar e separar os grãos de pólen.

B. Estudo da Taxa de Aprendizagem

Na busca da melhor taxa de aprendizado, observou-se que o valor de 10^{-3} foi adequado, sendo observado o resultado da experimentação da busca para melhor taxa de aprendizado na Figura 3.

Na literatura, em especial, na transferência de aprendizado, recomenda-se a diminuição da magnitude da taxa de aprendizado. No entanto, não foi constatada melhora no aprendizado no nosso pequeno conjunto de imagens. Houve um decaimento menos brusco na função da perda na validação com a taxa de aprendizado para gradiente estocástico descendente com 10^{-4} . No entanto, na precisão da validação os resultados se mostraram aquém dos restantes.

V. CONCLUSÃO

No presente artigo foi proposta uma nova abordagem com algoritmos de aprendizado profundo, baseando-se na técnica de ajuste fino para a classificação de grãos de pólen. Verificou-se que, dentro de algumas das principais arquiteturas convolucionais aquelas que são mais adequadas para tamanho do conjunto de dados. Como trabalhos futuros, pretende-se analisar os padrões aprendidos dentro da visualização das camadas ocultas, que podem fornecer uma intuição de como melhorar cada método abordado.

REFERÊNCIAS

[Boyain-Goitia et al., 2003] Boyain-Goitia, A. R., Beddows, D. C., Griffiths, B. C., and Telle, H. H. (2003). Single-pollen analysis by laser-induced breakdown spectroscopy and raman microscopy. *Applied Optics*, 42(30):6119–6132.

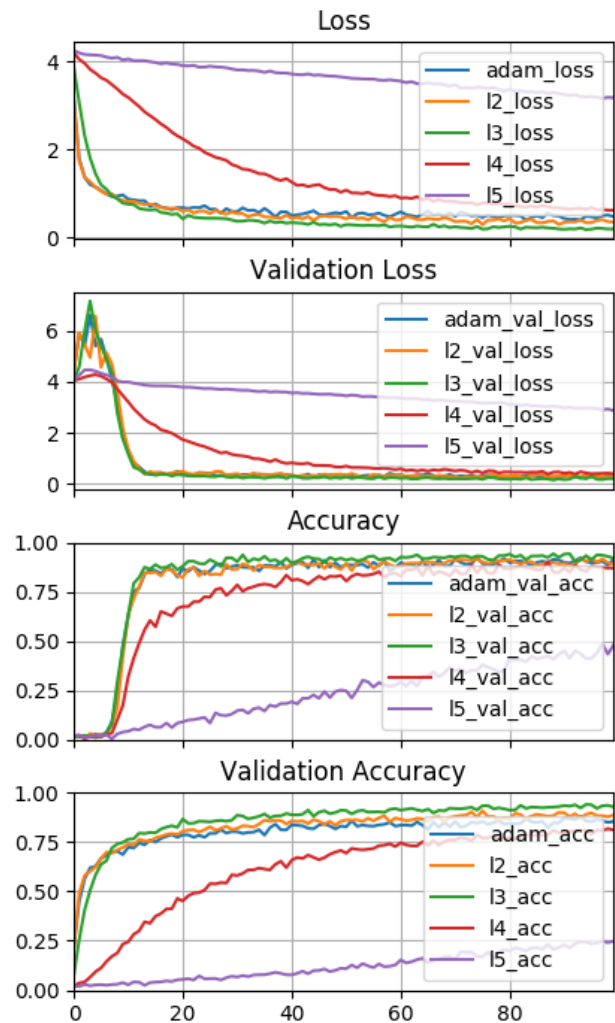


Figura 3. Estudo da Taxa de Aprendizagem para ResNet50, com 100 épocas

[Collinvaux et al., 2014] Collinvaux, P. A., De Oliveira, P. E., and Moreno, E. (2014). *Amazon: Pollen Manual and Atlas: Pollen Manual and Atlas*. CRC Press.

[Da Silva et al., 2014] Da Silva, D. S., Quinta, L. N. B., Goncalves, A., Pistori, H., and Borth, M. R. (2014). Application of wavelet transform in the classification of pollen grains. *African Journal of Agricultural Research*, 9(10):908–913.

[Delves et al., 2017] Delves, P. J., Martin, S. J., Burton, D. R., and Roitt, I. M. (2017). *Essential immunology*. John Wiley & Sons.

[Gonçalves et al., 2013] Gonçalves, A. B., Rodrigues, C. N. M., Cereda, M. P., and Pistori, H. (2013). Identificação computadorizada de tipos polínicos através de bag of words. *Cadernos de Agroecologia*, 8(2):1.

[Gonçalves et al., 2016] Gonçalves, A. B., Souza, J. S., da Silva, G. G., Cereda, M. P., Pott, A., Naka, M. H., and Pistori, H. (2016). Feature extraction and machine learning for the classification of brazilian savannah pollen grains. *PLoS one*, 11(6):e0157044.

[Hwang et al., 2013] Hwang, G. M., Riley, K. C., Christou, C. T., Jacyna, G. M., Woodard, J. P., Ryan, R. M., Bush, M. B., Valencia, B. G., McMichael, C. N., Punyasena, S. W., et al. (2013). Automated pollen identification system for forensic geo-historical location applications. In *2013 IEEE International Conference on Technologies for Homeland Security (HST)*, pages 297–303. IEEE.

[Jackson and Williams, 2004] Jackson, S. T. and Williams, J. W. (2004). Modern analogs in quaternary paleoecology: here today, gone yesterday, gone tomorrow? *Annu. Rev. Earth Planet. Sci.*, 32:495–537.

- [Joosten and de Klerk, 2002] Joosten, H. and de Klerk, P. (2002). What's in a name?: Some thoughts on pollen classification, identification, and nomenclature in quaternary palynology. *Review of Palaeobotany and Palynology*, 122(1-2):29–45.
- [Kingma and Ba, 2014] Kingma, D. P. and Ba, J. (2014). Adam: A method for stochastic optimization. *arXiv preprint arXiv:1412.6980*.
- [Nakagawa et al., 2000] Nakagawa, T., Edouard, J.-L., and de Beaulieu, J.-L. (2000). A scanning electron microscopy (sem) study of sediments from lake cristol, southern french alps, with special reference to the identification of pinus cembra and other alpine pinus species based on sem pollen morphology. *Review of Palaeobotany and Palynology*, 108(1-2):1–15.
- [Ng, 2019] Ng, A. (2019). Machine learning yearning: Technical strategy for ai engineers in the era of deep learning. Technical report, Tech. Rep.
- [Prince, 2012] Prince, S. J. (2012). *Computer vision: models, learning, and inference*. Cambridge University Press.
- [Quinta, 2013] Quinta, L. N. B. (2013). Visão computacional aplicada na classificação de grãos de pólen. Master's thesis, Programa de Pós Graduação em Biotecnologia, Universidade Católica Dom Bosco.
- [Quinta et al., 2012] Quinta, L. N. B., Amorim, W. P., de Carvalho, M. H., Cereda, M. P., and Pistori, H. (2012). Floresta de caminhos otimos na classificaç ao de pólen. In *Workshop de Visão Computacional*.
- [Russakovsky et al., 2015] Russakovsky, O., Deng, J., Su, H., Krause, J., Satheesh, S., Ma, S., Huang, Z., Karpathy, A., Khosla, A., Bernstein, M., et al. (2015). Imagenet large scale visual recognition challenge. *International journal of computer vision*, 115(3):211–252.
- [Von Der Ohe et al., 2004] Von Der Ohe, W., Oddo, L., Piana, M., Morlot, M., and Martin, P. (2004). Harmonized methods of melissopalynology. 35:18–25.