



Data Augmentation via Aprendizado Semissupervisionado Baseado em Grafos

Willian Dihanster Gomes de Oliveira

Orientadora: Lilian Berton, Co-orientador: Otávio A. B. Penatti

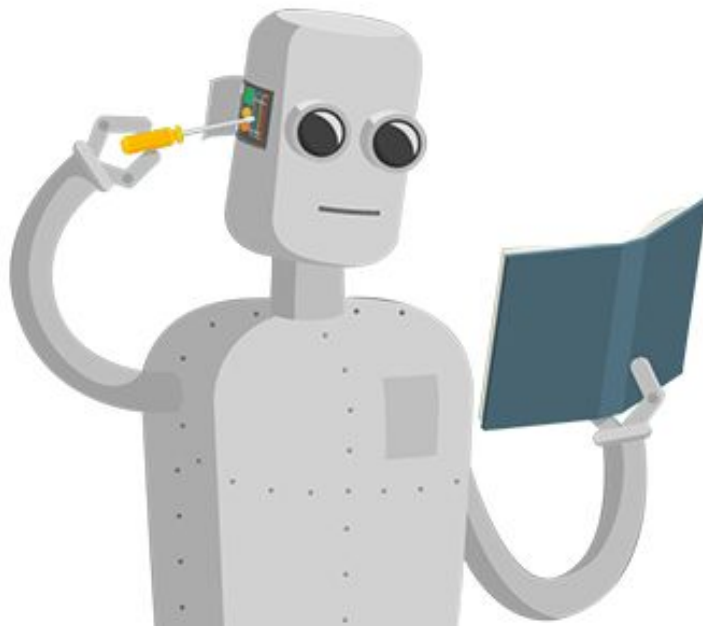
Instituto de Ciência e Tecnologia, Universidade Federal de São Paulo

11 de Junho de 2018

São José dos Campos - SP

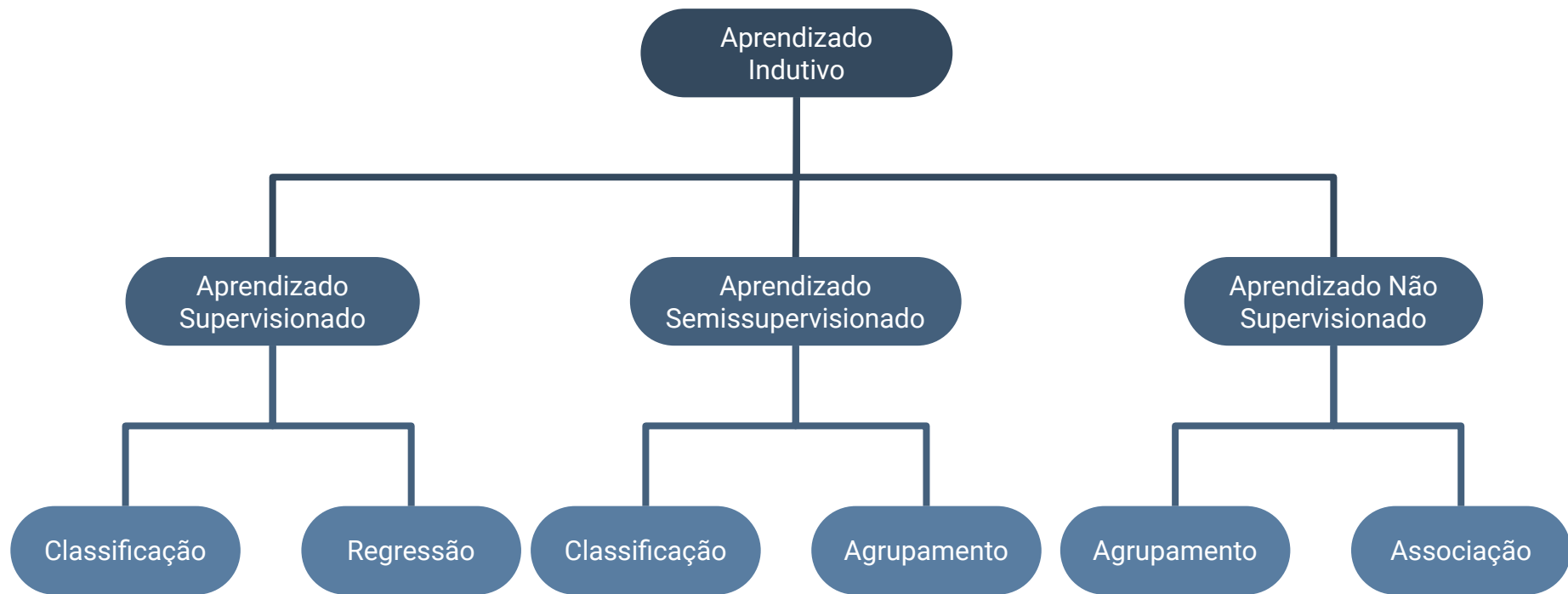
Introdução

- Aprendizado de Máquina (AM):
 - aprender a partir de exemplos.
 - melhorar automaticamente com a experiência. [1]



Robô ajustando seus parafusos para aprender. Disponível em:
<http://fiori-implementation.com/blog/make-clean-breast-machine-learning/>

Hierarquia do Aprendizado



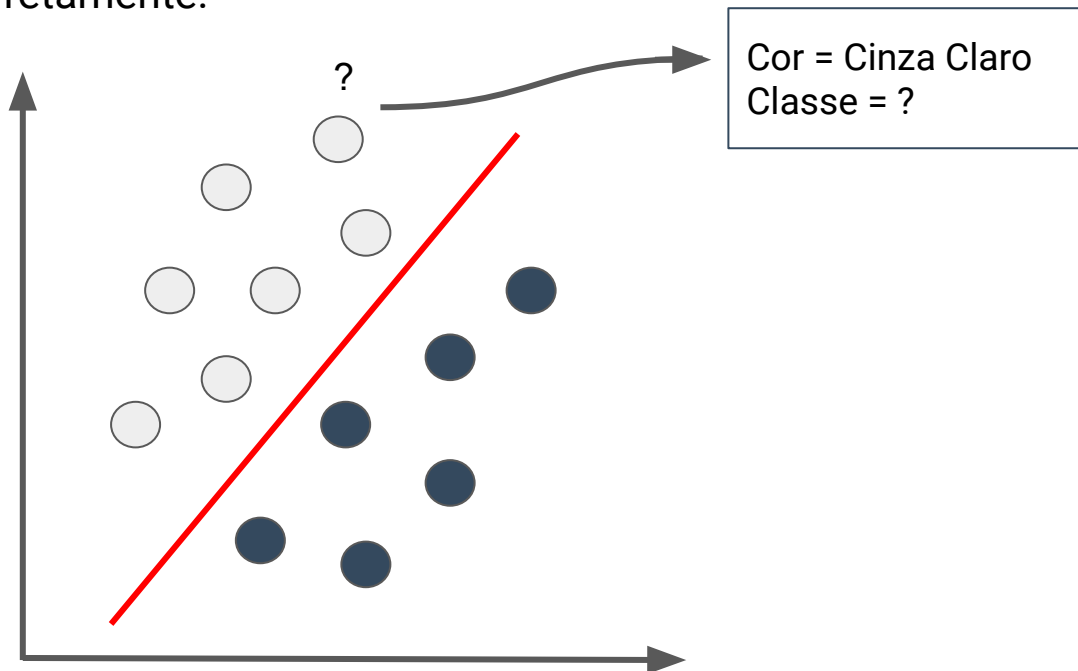
Classificação

- Classe:
 - atributo especial que rotula um elemento, atributo que queremos prever.
- O objetivo da classificação é modelar uma função que para cada elemento ainda não visto, consiga prever seu rótulo corretamente.

Cor = Cinza Claro
Classe = 0

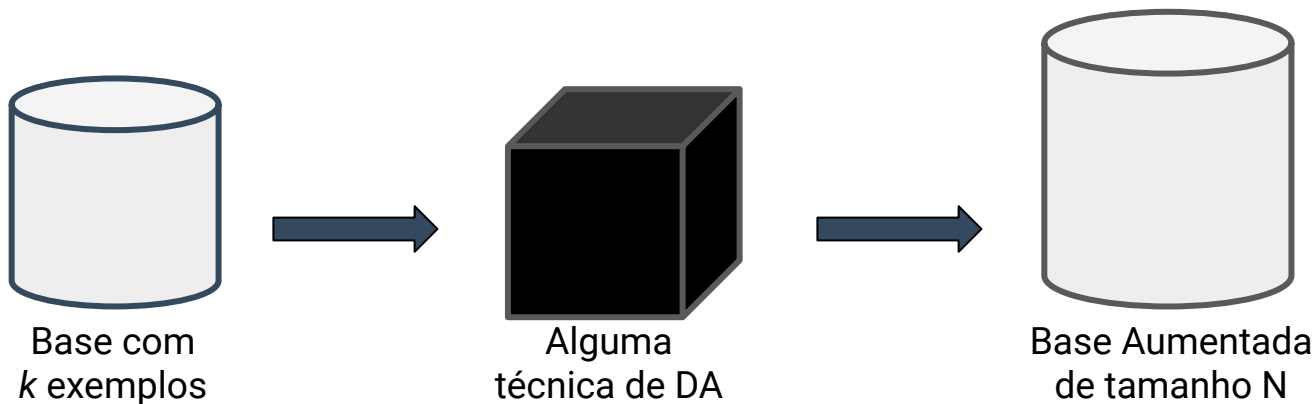


Cor = Azul Forte
Classe = 1



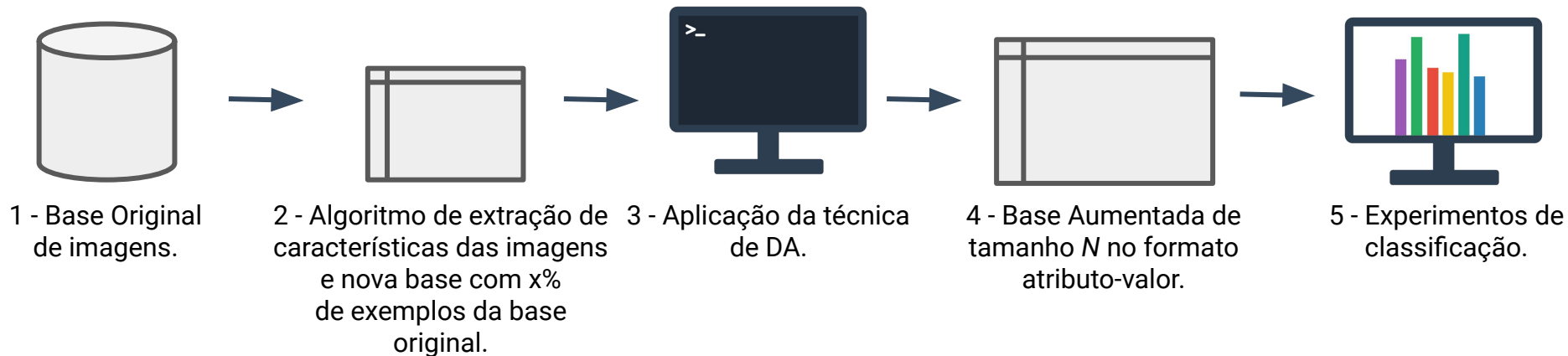
Data Augmentation (DA)

- Aprendizado Supervisionado: quanto mais dados (rotulados), maior a acurácia.
- Mas, dados rotulados são custosos e difíceis de se obter.
- *Data Augmentation*:
 - Técnicas computacionais para aumentar bases de dados.
- Objetivo: estudo de duas técnicas de DA para a classificação de imagens.



Metodologia

- A metodologia é descrita como a sequência de passos a seguir:

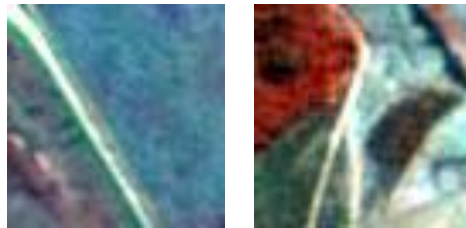


Base de Dados

- A base de dados utilizada (Brazilian Coffee Scenes Dataset [2]) é composta por imagens de satélite de plantações de café, em cidades de Minas Gerais pelo sensor SPOT.
- Cada imagem possui 64x64 pixels e foram classificadas manualmente por pesquisadores agrícolas como "coffee" e "non coffee".
- 2876 imagens, 50% de coffee e 50% non coffee.



Exemplos rotulados como coffee.



Exemplos rotulados como non coffee.

Técnica de DA 1: Transformação de Imagens (Imagens)

- Nessa abordagem, são escolhidas algumas imagens aleatoriamente do conjunto original e aplica-se alguns efeitos como corte, girar, *blur*, até expandir para um novo conjunto de tamanho N .



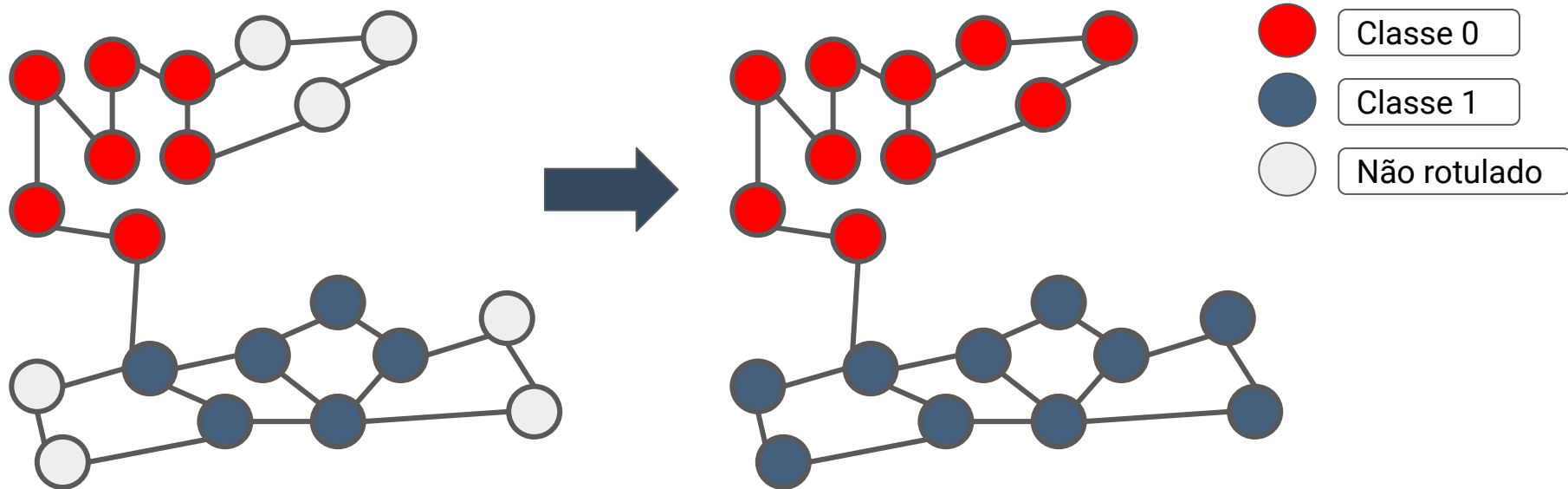
Exemplo aleatório
rotulado como coffee.



Exemplo de coffee depois de
aplicado efeito *blur* e girar.

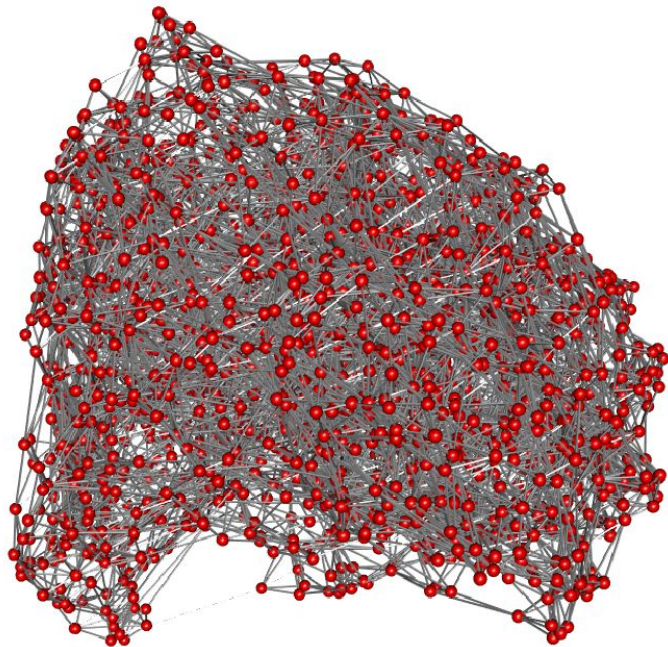
Técnica de DA 2: Aprendizado Semissupervisionado Baseado em Grafos

- Elementos rotulados e não rotulados. Pode-se tentar classificar esses dados não rotulados.
- Algoritmos de Propagação de Rótulos:
 - A partir de dados rotulados, propagar sua informação de classe para os dados ainda não rotulados.
 - Para isso pode-se usar um grafo.



Aprendizado Semissupervisionado Baseado em Grafos (Grafos-SSL)

- Constrói-se um grafo com a informação de semelhança dos exemplos e é aplicado o algoritmo de propagação de rótulos *Local and Global Consistency (LGC)* [3].
- Pontos próximos (local) e de mesma estrutura (global) são prováveis de terem mesmo rótulo.



Grafo construído para a propagação.

Algoritmo 1: Local and Global Consistency

Entrada: G um grafo ponderado $G(V, E, W)$.

Saída: Rótulos para o conjunto Y_u , $l + 1 \leq u \leq n$

início

Calcula-se a matriz de afinidade W , utilizando a seguinte fórmula:

$$W_{ij} = \begin{cases} \exp(-\frac{\|x_i - x_j\|^2}{2\sigma^2}) & \text{se } i \neq j \\ 0 & \text{se } i = j \end{cases} \quad (19)$$

Calcula-se a matriz diagonal D , como: $D_{ii} \leftarrow \sum_j W_{ij}$.

Calcula-se a matriz laplaciana normalizada utilizando: $S = D^{1/2} W D^{1/2}$.

Escolhe-se um parâmetro α no intervalo $(0, 1)$.

repita

$F(t + 1) = \alpha S F(t) + (1 - \alpha) Y$.

até que F atinja a convergência;

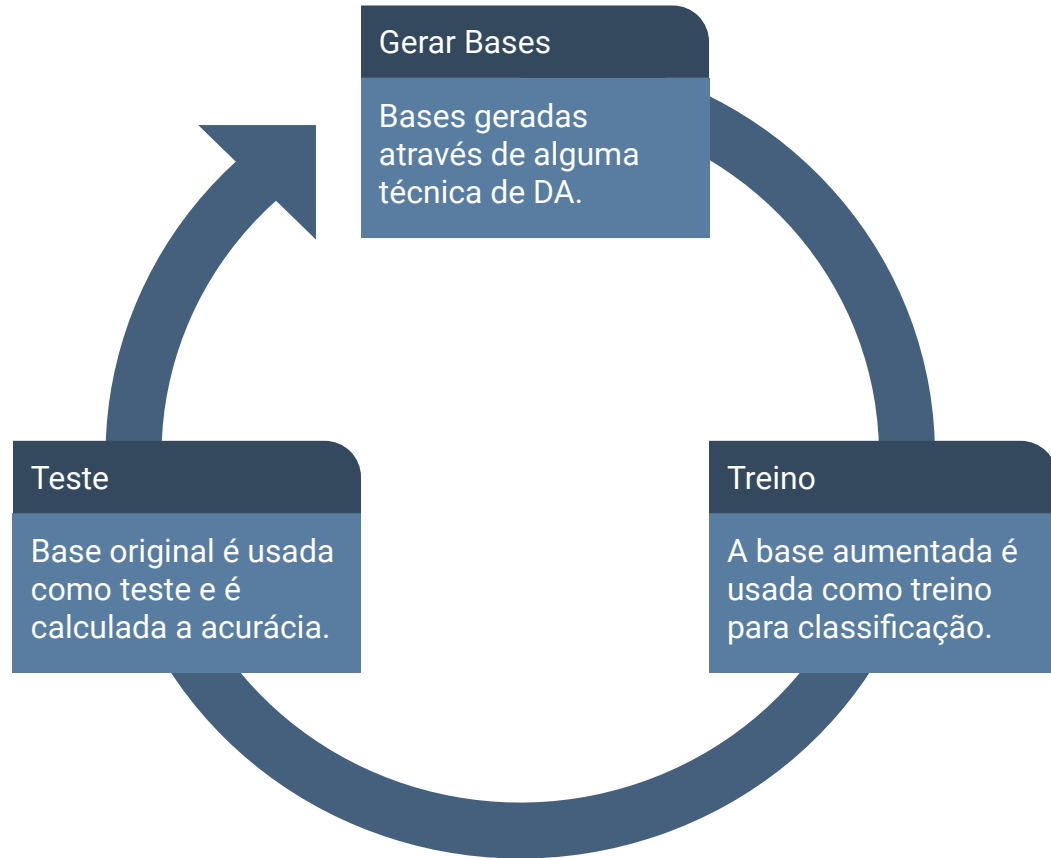
Ao final, obtém-se F^* .

Rotula-se as amostras x_i , com $y_i = \arg \max_{j \leq c} F_{ij}^*$.

fim

Algoritmo LGC.

Experimentos



Softwares e classificadores utilizados para os experimentos

- Softwares utilizados:
 - WEKA
 - Tensorflow
- Classificadores/Algoritmos utilizados (parâmetros padrão):
 - Naive Bayes
 - IBK (kNN = k -Nearest Neighbors)
 - $k = 1, k = 3$ e $k = 5$;
 - MLP (Multilayer Perceptron - Redes Neurais)
 - SMO (SVM - Support Vector Machine)
 - J48 (Árvores de Decisão)

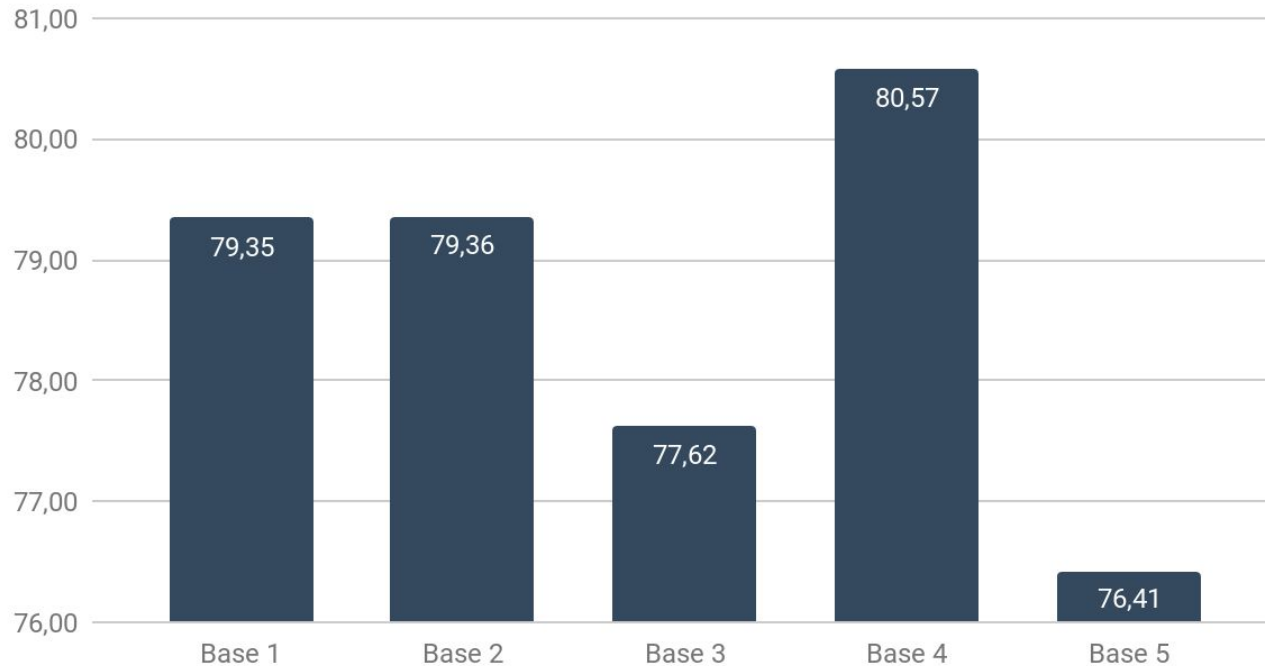
Experimentos Iniciais

- Para os experimentos iniciais foram geradas 15 bases de dados:
 - Ambos a partir de 10% do conjunto original.
 - 5 bases usando transformação de imagens:
 - cada base possui 2876 exemplos.
 - 10 bases usando aprendizado semissupervisionado:
 - cada base possui 2876 exemplos.
 - Software utilizado: WEKA

Resultados

- Imagens

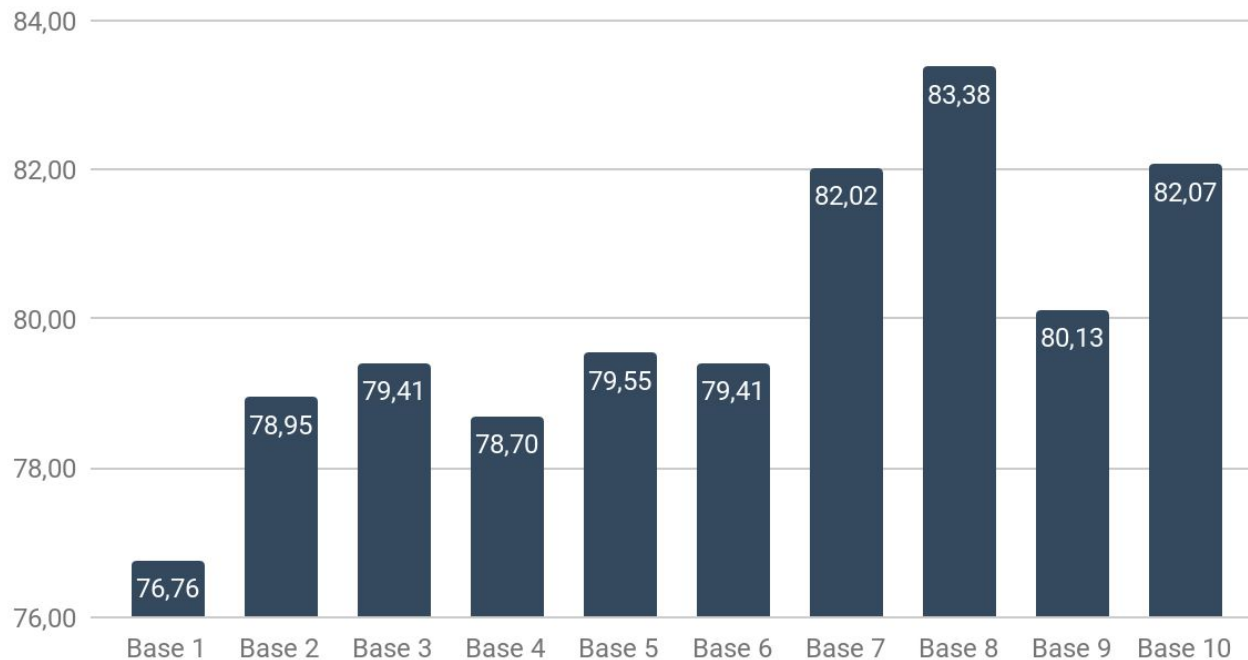
Média dos Resultados no WEKA (Imagens)



Resultados

- Grafos-SSL

Média dos Resultados no WEKA (Grafos-SSL)



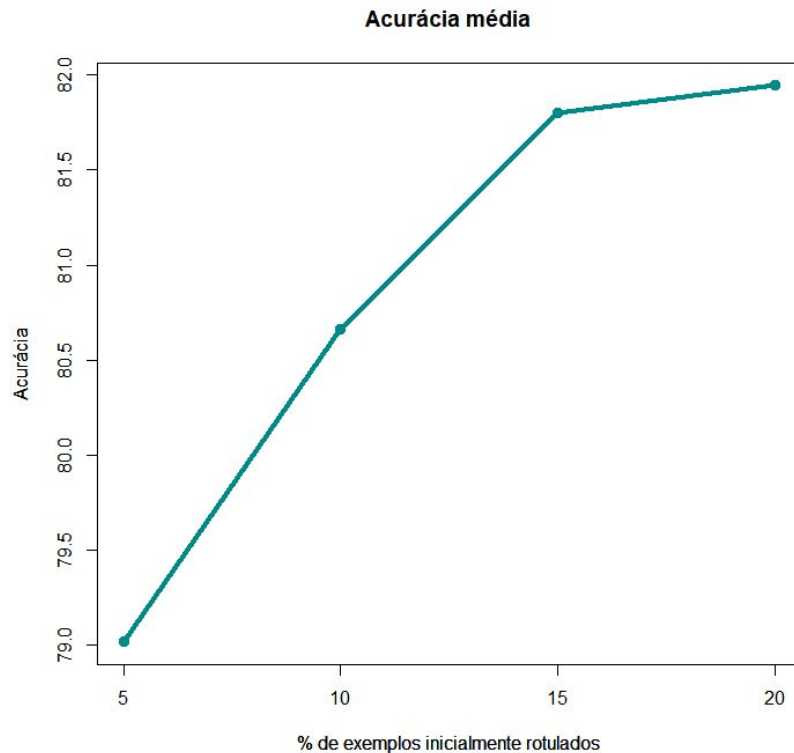
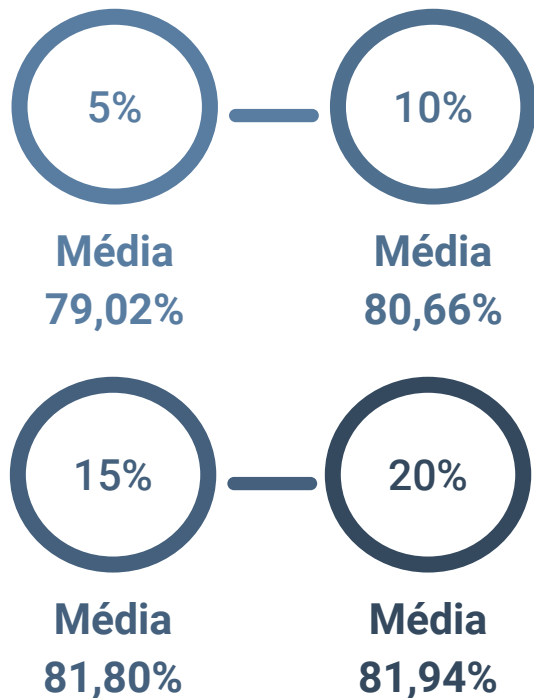
Resultados

- Transformação de Imagens x Grafos-SSL

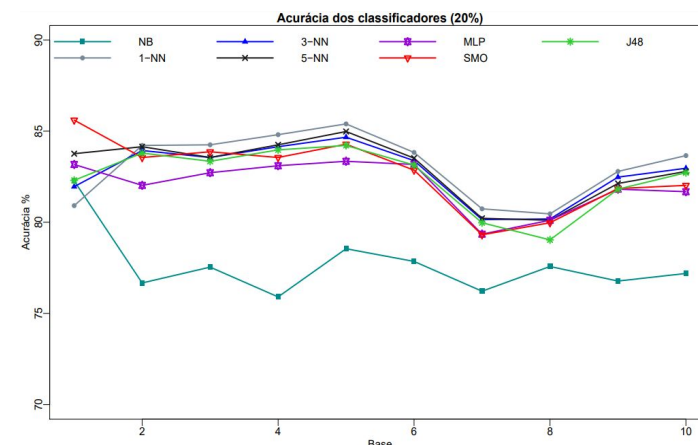
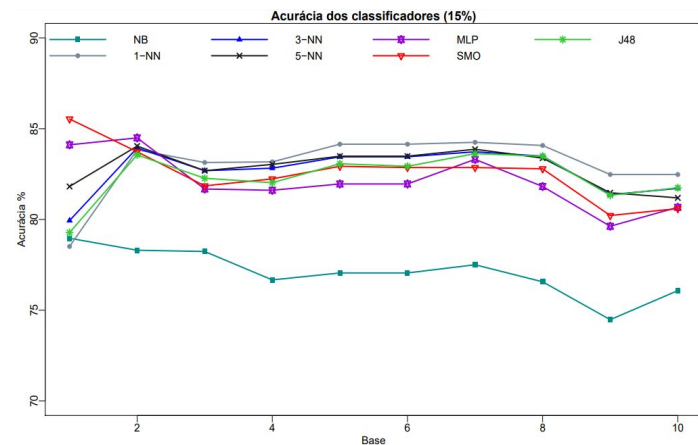
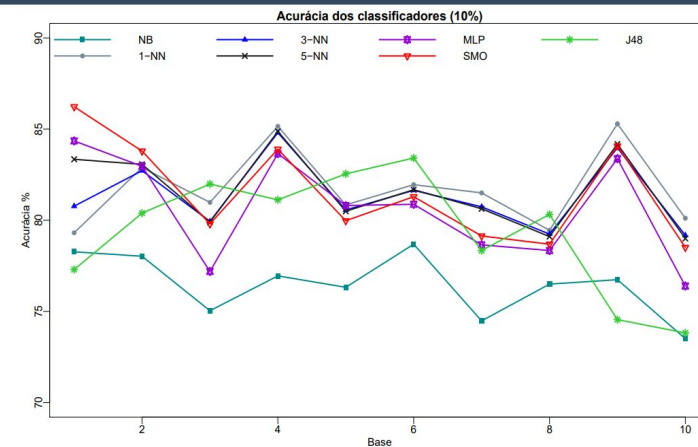
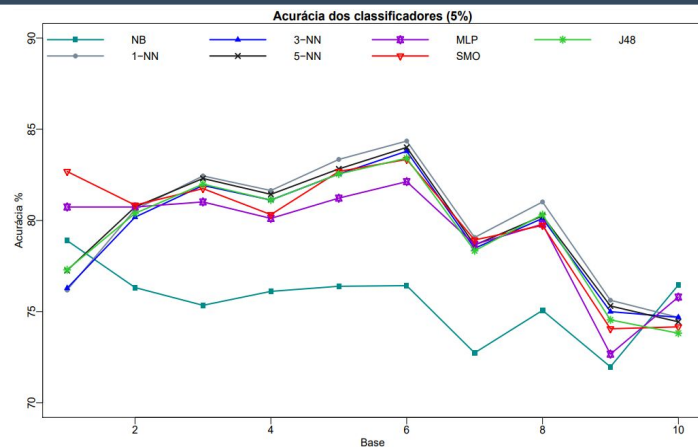
	Média	Desvio Padrão	Máximo	Mínimo
Imagens	78,66%	1,64	82,68%	70,97%
Grafos-SSL	80,04%	1,94	84,98%	73,61%
Original (10-fold)	84,37%	2,65	87,41%	79,73%

Grafos-SSL com variação na % inicial

- 40 novas bases utilizando Grafos-SSL
 - 10 bases para cada % de exemplos inicialmente rotulados.



Comparação dos Algoritmos utilizados para cada % inicial no WEKA



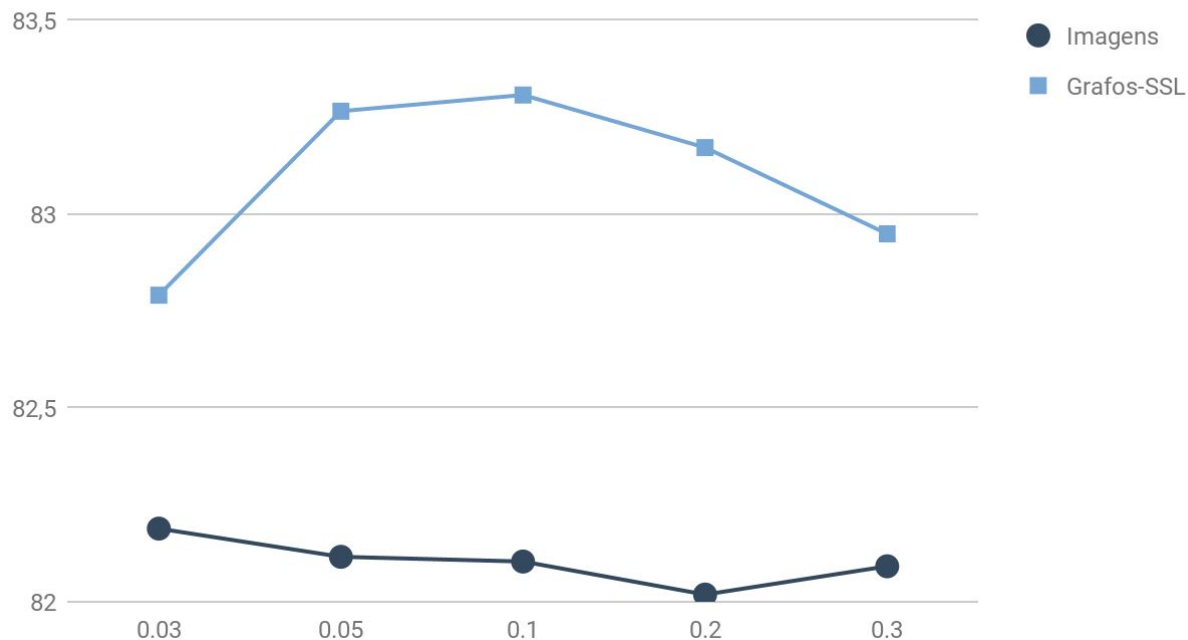
Deep Learning e Variação de Parâmetros

- Foco nas redes neurais profundas.
 - Escolheu-se a base com maior desempenho médio de Imagens e de Grafos-SSL.
- Varia-se:
 - *Learning Rate* (Taxa de Aprendizado).
 - Número de Neurônios em cada camada.
 - Tempo de Treinamento (épocas).
- Todas essas variações são consideradas para:
 - 3 camadas ocultas no WEKA.
 - 3, 4 e 5 camadas ocultas no Tensorflow.

Learning Rate (Taxa de Aprendizado)

- Variações (0.03, 0.05, 0.1, 0.2 e 0.3). Resultado médio para as 4 redes:

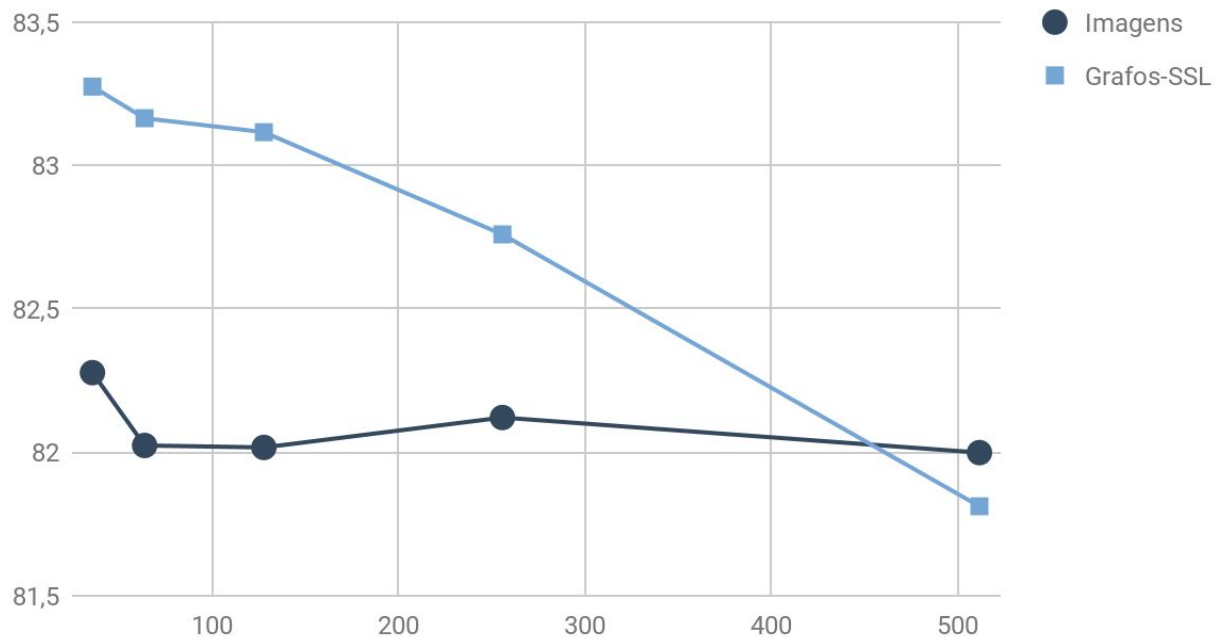
Variação no Learning Rate



Número de Neurônios nas camadas ocultas

- Número de Neurônios (32, 64, 128, 256 e 512). Resultado médio para as 4 redes:

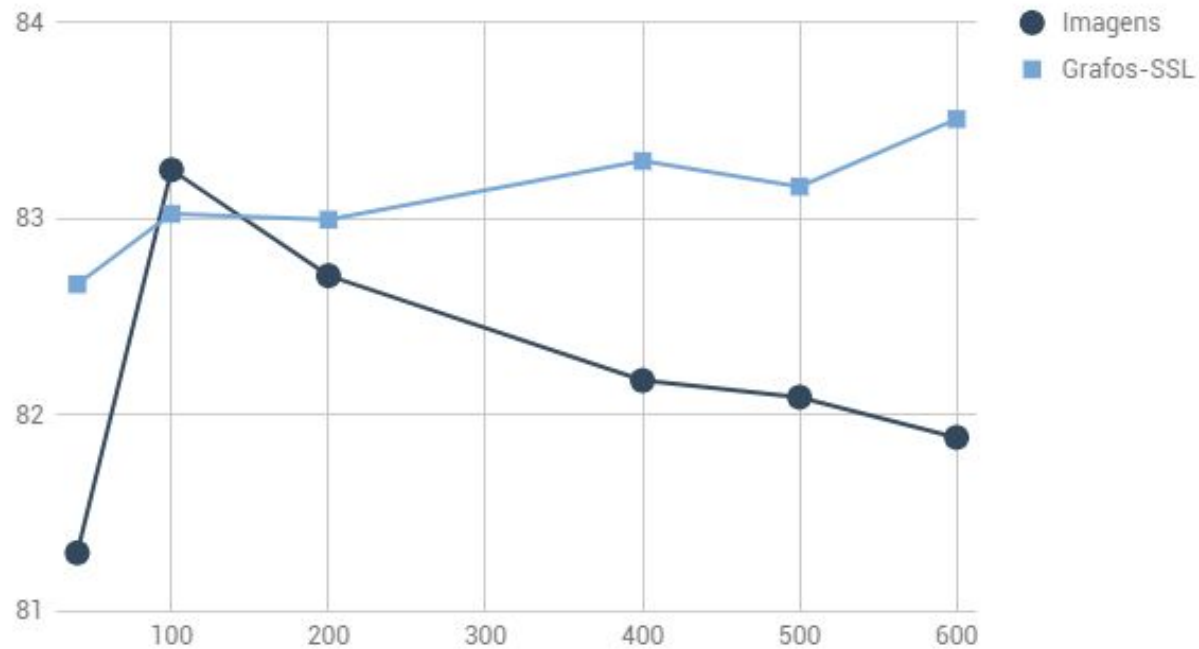
Variação no número de neurônios em cada camada oculta



Tempo de Treinamento (Épocas)

- Épocas (40, 100, 200, 400, 500, 600). Resultado médio para as 4 redes:

Variação do Tempo de Treinamento



Resumo Variação de Parâmetros para o WEKA

- Melhor resultado no WEKA para Imagens foi **84,46%** e Grafos-SSL **84,11%** com os parâmetros descritos pela figura a seguir.

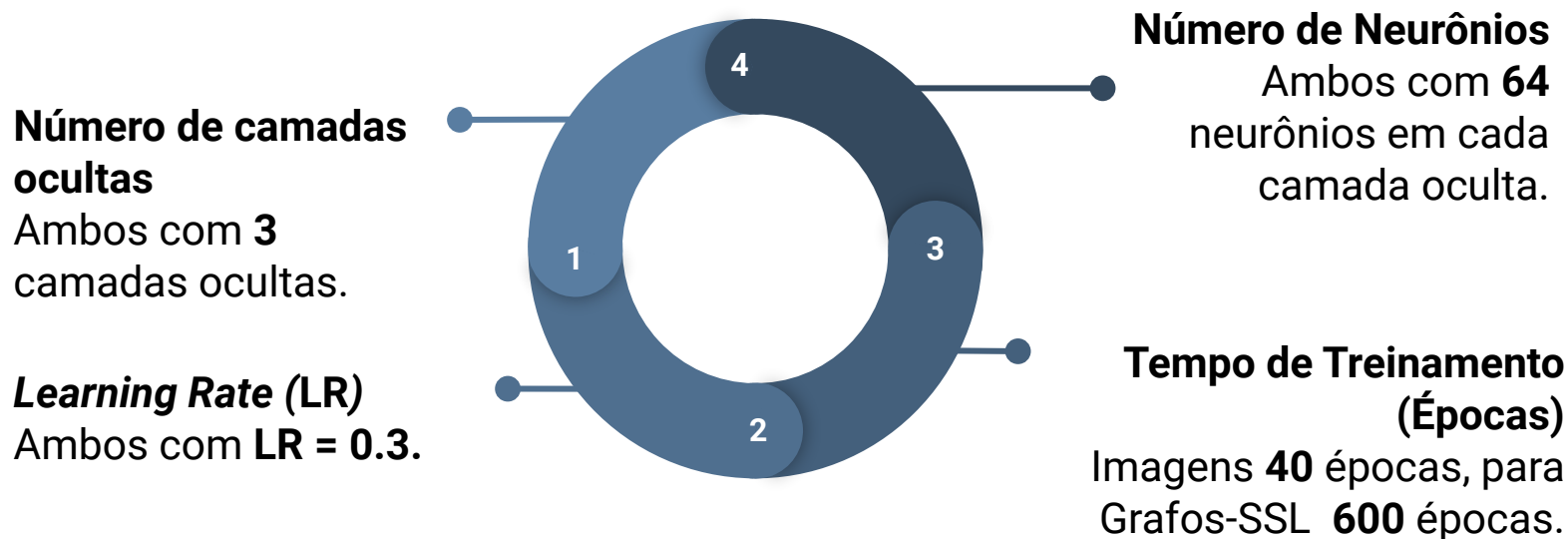


Diagrama dos parâmetros dos melhores resultados encontrados.

Resumo Variação de Parâmetros para o Tensorflow

- Melhor resultado no Tensorflow para Imagens foi **84,28%** e Grafos-SSL **83,92%** com os parâmetros descritos pela figura a seguir.

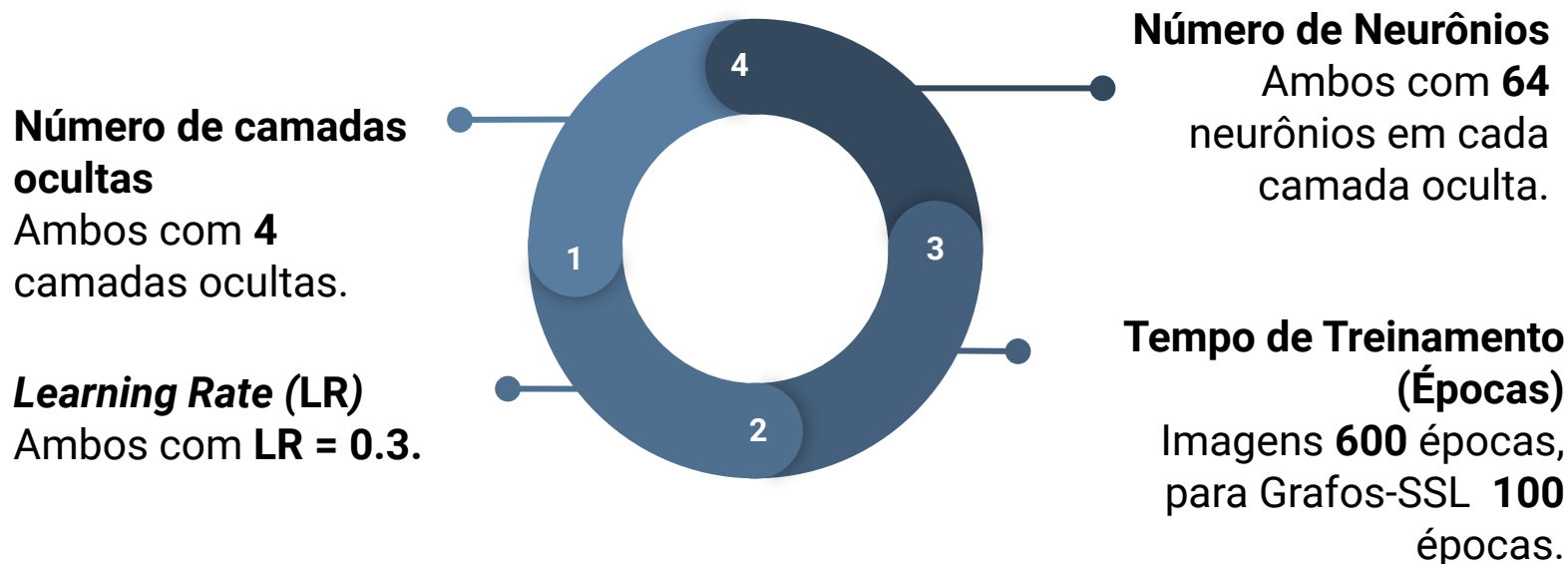


Diagrama dos parâmetros dos melhores resultados encontrados.

Conclusão

- Ambas técnicas podem ser consideradas eficientes para a aplicação escolhida. E que DA pode melhorar a acurácia no caso de poucos exemplos rotulados.
- Abordagem Grafos-SSL no geral, melhores resultados em relação à técnica de imagens.
- Grafos-SSL
 - Vantagens: Amplo, pode se aplicar diversos algoritmos de propagação, alterações na construção do grafo. Melhores resultados (para a aplicação escolhida).
 - Desvantagens: Somente em abordagens em que se possuem elementos rotulados e não rotulados.
- Transformação de Imagens:
 - Vantagens: Pode-ser aplicado a qualquer conjunto de dados de imagens.
 - Desvantagens: Cuidado com a redundância ou perda de sentido nos dados.

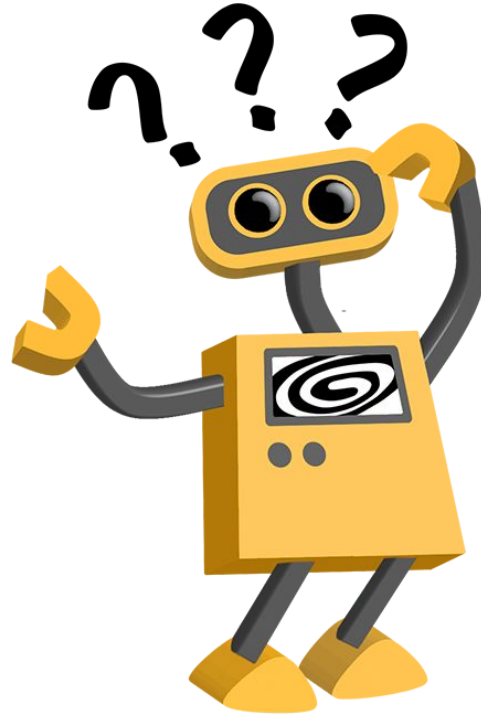
Agradecimentos

- Esse projeto foi apoiado e financiado pela Universidade Federal de São Paulo.



Alguma Dúvida ?

- Obrigado pela atenção!



Disponível em: <https://fcit.usf.edu/matrix/project/robot-10-robot-with-questions/>

Referências

- [1] Michalski, Ryszard S., et al., eds. *Machine learning: A multistrategy approach*. Vol. 4. Morgan Kaufmann, 1994.
- [2] Penatti, Otávio AB, Keiller Nogueira, and Jefersson A. dos Santos. "Do deep features generalize from everyday objects to remote sensing and aerial scenes domains?." *Computer Vision and Pattern Recognition Workshops (CVPRW), 2015 IEEE Conference on*. IEEE, 2015.
- [3] Zhou, Denny, et al. "Learning with local and global consistency." *Advances in neural information processing systems*. 2004.