

Uma Abordagem de Aprendizado Profundo e *Data Augmentation* para Classificação de Imagens Aéreas

Willian Dihanster Gomes de Oliveira, Lilian Berton
Instituto de Ciência e Tecnologia - Universidade Federal de São Paulo

Resumo—O Aprendizado de Máquina (AM) é uma área de grande destaque nos dias atuais. No entanto, um dos seus problemas, é a necessidade de uma grande quantidade de dados, especialmente rotulados, que são custosos e difíceis de obter. Assim, várias técnicas têm sido propostas para obter dados mais rotulados. Neste trabalho, objetivamos estudar duas técnicas de *Data Augmentation* (DA) para classificação de imagens. Além disso, uma subárea de grande destaque no AM é o *Deep Learning* que vem obtendo grandes resultados, especialmente para classificação de imagens. Assim, aplicamos técnicas de DA em conjunto com o *Deep Learning* para classificar imagens aéreas. Os resultados indicam que ambas as técnicas alcançaram bons resultados.

I. INTRODUÇÃO

O Aprendizado de Máquina (AM) é uma área de grande destaque nos dias atuais, a qual se baseia em criar e modelar sistemas capazes de aprender a partir de exemplos e que melhorem automaticamente com a experiência. No AM há uma hierarquia do aprendizado que pode ser dividida entre: aprendizado supervisionado, semissupervisionado e não supervisionado. [1] No aprendizado supervisionado, onde é comum a classificação, há uma grande variedade de algoritmos, como *k*-vizinhos mais próximos (*k*NN), Árvores de Decisão, *Support Vector Machines*, Redes Neurais, etc. Atualmente, um classificador em destaque são as Redes Neurais Profundas, que vêm obtendo ótimos resultados na classificação de imagens, pois possuem uma alta capacidade de abstração. [2]

Dentre as maiores aplicações da classificação de imagens estão a automatização de tarefas. Um exemplo é a classificação de imagens de plantações de café que podem ajudar os agricultores a identificar mais facilmente, lugares de sua plantação que possuem um pé de café, com a ajuda de um *software* e um sensor, sem a necessidade de percorrer toda sua plantação. [3] Entretanto, para que esse reconhecimento automático funcione da melhor maneira possível é necessário uma grande quantidade de dados e em especial, rotulados para treinar um classificador. O que é uma tarefa difícil e custosa de se fazer. Assim, diversas técnicas de *Data Augmentation* (DA) vêm sendo empregadas na obtenção de mais dados rotulados.

Neste trabalho, objetivamos o estudo de duas técnicas de DA aplicada a classificação de imagens de satélite de plantações de café. A técnica de Transformação de Imagens (será denotado como "Imagens") expande o conjunto de dados aplicando efeitos como corte, girar, *blur*, etc. E a técnica de Aprendizado Semissupervisionado Baseado em Grafos (denotado por "Grafos-SSL- *Semi-supervised Learning*") expande o conjunto de dados propagando a informação de

classe de alguns dados rotulados para os elementos não rotulados.

Os experimentos de classificação foram realizados utilizando as Redes Neurais Profundas com a biblioteca TensorFlow para Python e as Redes Neurais com 3 camadas ocultas no WEKA. Foram feitas variações no número de camadas ocultas e alguns dos principais parâmetros. Foi analisado os resultados de acurácia na classificação empregando ambas as técnicas de DA, bem como as vantagens e desvantagens de cada técnica.

O restante do artigo está organizado como segue: Na Seção II será detalhado alguns conceitos sobre a classificação, com foco em redes neurais, *Deep Learning* e *Data Augmentation*. Na Seção III são sumarizados alguns trabalhos relacionados. Na Seção IV é descrita a metodologia empregada. Na Seção V é apresentada a análise experimental e por fim, na Seção VI é apresentada as conclusões do trabalho.

II. CLASSIFICAÇÃO

Na classificação de dados, os algoritmos recebem exemplos rotulados com um atributo classe e aprendem comparando a saída gerada pelo algoritmo com a saída correta. A partir dos erros identificados modificam o modelo para se ajustar as saídas corretas. [4] O objetivo do indutor é gerar um classificador que sirva para rotular novos elementos que não faziam parte da amostra inicial, usada no treinamento.

Há uma diversa variedade de algoritmos de classificação disponíveis em AM, o algoritmo Redes Neurais utilizado nesse trabalho é detalhado na seção A. Enquanto que o problema da falta de dados e uma possível solução é comentada na seção B.

A. Redes Neurais (RN) e Deep Learning

Um dos classificadores mais populares em AM são as Redes Neurais. As Redes Neurais são baseadas em um modelo matemático de como o cérebro realiza suas tarefas e como aprende. O *Multilayer Perceptron* (MLP) simples é um modelo de Rede Neural que até então, era o mais utilizado. Nesse modelo, em geral, há a presença de uma camada de entrada, camada de saída e uma camada oculta. [5]

Com o crescimento do poder de processamento dos computadores atuais, houve a ascensão do *Deep Learning* ou Aprendizado Profundo. Essa forma de aprendizado vem obtendo ótimos resultados, pois permitem uma maior capacidade de abstração dos problemas. [2] No *Deep Learning* há também

uma grande variedade de modelos de Redes Neurais, por exemplo, o modelo mais simples é basicamente uma MLP com mais camadas ocultas, as Redes Neurais Convolucionais (possuem camadas de convolução), dentre outras.

Para esse trabalho, foi considerado o modelo mais simples, uma MLP com 3 e 5 camadas ocultas e uma MLP de 4 camadas ocultas com regularizadores detalhados a seguir.

1) *Multilayer Perceptron*: Esse modelo de rede neural profunda pode ser vista como uma Rede MLP padrão com mais camadas ocultas. No modelo utilizado, os neurônios são *fully connected*, isto é, todos os neurônios se ligam a todos os outros das camadas anterior e seguinte. Nesse trabalho foi utilizado MLPs com 3 e 5 camadas ocultas.

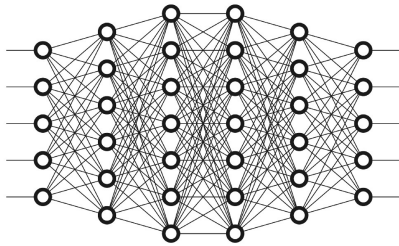


Figura 1: Exemplo de uma Rede MLP Profunda.

Fonte: Alex Woodie. "Neural networks can have many hidden layers", Datanami 2017

2) *Multilayer Perceptron (com Dropout)*: Esse modelo é parecido com a rede MLP, com a diferença de aplicação das técnicas de regularização, como o *Dropout*. No *Dropout*, alguns neurônios são desativados temporariamente com uma probabilidade p . Essas técnicas de regularização reduzem a chance de *overfitting* [6] adicionando uma penalidade para a função de perda. Com essa adição, há uma redução da interdependência de aprendizado entre os neurônios. Algumas das técnicas são a Regressão Logística com penalidades L1 (Laplaciano) e L2 (Gaussiana). Nesse caso, usou-se 4 camadas ocultas.

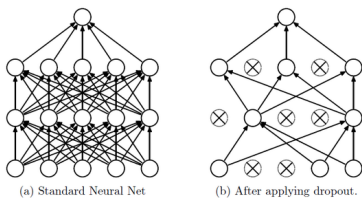


Figura 2: Exemplo de uma Rede MLP Profunda com Dropout.

Fonte: Srivastava, Nitish, et al. "Dropout: a simple way to prevent neural networks from overfitting", JMLR 2014

B. Data Augmentation (DA)

Em AM, o desempenho dos algoritmos no aprendizado supervisionado pode ser estendido de acordo com o aumento no número de instâncias numa base de dados. No entanto, em situações da vida real, nem sempre temos muitos exemplos disponíveis ou muitas vezes, não são rotulados.

Dessa forma, criou-se o conceito de *Data Augmentation* (DA), que em tradução direta, significa aumento de dados. O DA tem como objetivo utilizar técnicas computacionais para aumentar a base de dados original e como consequência, obter uma melhor acurácia na classificação. Duas técnicas de DA são descritas a seguir.

1) *Transformações de Imagens*: Quando se trabalha com imagens e deseja-se aumentar essa base, uma abordagem simples e muito utilizada é a de transformação de imagens. Nessa técnica, a partir de um conjunto de imagens determinadas são aplicadas edições de imagem como corte, rotação, *blur*, na imagem original, gerando uma nova imagem. Assim, é gerado uma base aumentada, composta pelas imagens originais mais as imagens editadas.

2) *Aprendizado Semissupervisionado Baseado em Grafos*: No aprendizado semissupervisionado, o qual considera elementos rotulados e não-rotulados, uma abordagem comum para aumentar o conjunto de dados é classificar os elementos não rotulados usando alguma técnica ou heurística. Assim, pode-se usar um algoritmo de propagação de rótulos, em que a ideia é usar os dados rotulados para propagar a informação de rótulo para os elementos que ainda não foram rotulados, os quais podem ser modelados através de um grafo.

III. TRABALHOS RELACIONADOS

A. Deep Features Generalize from Everyday Objects to Remote Sensing and Aerial Scenes Domains?

A ideia dos autores [3] era estimar o poder de generalização das *Convolutional Neural Networks* (CNNs) aplicado na classificação de imagens de sensores remotos e imagens aéreas. O domínio utilizado é o mesmo domínio de imagens de plantações de café do trabalho aqui realizado. No entanto, no nosso trabalho, utilizamos técnicas de DA para as bases utilizadas, o que tenta aproximar os resultados obtidos da base completa.

B. Research on Data Augmentation for Image Classification Based on Convolutional Neural Networks

A proposta dos autores [7] era fazer DA a partir de transformação de imagens, utilizando técnicas de girar, rotacionar, cortar, etc e/ou uma combinação desses métodos. Após a geração da base aumentada, classificar os dados utilizando as CNNs e tentar descobrir dentre essas técnicas de transformação de imagens, quais são as melhores dentre os conjuntos de dados.

C. Effectiveness of Data Augmentation in Image Classification using Deep Learning

Neste trabalho os autores [8] propuseram o estudo de outras técnicas de DA para expandir o conjunto de imagens. Nesse caso, além das técnicas tradicionais como corte, girar, etc, consideraram também filtros de estilização das imagens. O domínio do problema foi baseado em imagens de animais, como peixes, cachorros, gatos e dígitos.

D. Operational Data Augmentation in Classifying Single Aerial Images of Animals

Baseado em técnicas de DA para classificação de imagens aéreas. A ideia dos autores [9] foi aplicar técnicas de transformação de imagens na base de dados original, com o diferencial de usar uma técnica que cria uma nova imagem a partir de diversos ângulos de rotação de uma só imagem. Isto é, forma uma espécie de grade com a mesma imagem, em vários ângulos de rotação. A aplicação é parecida com a do nosso trabalho por se tratar de imagens aéreas, mas numa situação mais simples, onde as classes para diferenciação são com presença de vaca ou sem presença de vaca.

E. Iterative Label Propagation on Facial Images

Com a ideia de propagação de rótulos para DA, a proposta dos autores [10] era levar em consideração as informações dos dados, utilizando clusterização para construir o grafo. Essa informação então é usada para regular a força de similaridade entre os dados e qual dado será selecionado para inicializar a propagação.

IV. MATERIAIS E MÉTODOS

Nessa seção é descrito a metodologia abordada para realização dos testes e as ferramentas utilizadas.

A. Base de dados

A base de dados original utilizada (*Brazilian Coffee Scenes Dataset* [3]) é composta por imagens de satélite de plantações de café, em cidades de Minas Gerais coletadas pelo sensor SPOT.

Cada uma das imagens possuem 64x64 pixels e foram classificadas manualmente por pesquisadores agrícolas como “coffee” e “non coffee”. Assim, o conjunto de dados é composto por 2876 exemplos, sendo 50% exemplos da classe “coffee” e 50% da classe “non coffee”. Nas Figuras 3 e 4, temos alguns exemplos das imagens capturadas pelo sensor.



Figura 3: Exemplo de imagens rotuladas como “coffee”. Fonte: Rubens Lamparelli, Cooxupé “Brazilian Coffee Scenes Dataset”.

B. Expansão das Bases - Data Augmentation

Para a expansão dos conjuntos de dados rotulados utilizamos duas abordagens:

1) *Transformações de Imagens*: A primeira delas faz uso de Transformações de Imagens. Considerando uma porcentagem inicial (10% de dados com rótulos) aplica-se diversas transformações como rotação, *blur*, etc nesse subconjunto até expandi-lo para um tamanho N . Na Figura 5 temos alguns exemplos das transformações aplicadas nas imagens da base original.



Figura 4: Exemplo de imagens rotuladas como “non-coffee”. Fonte: Rubens Lamparelli, Cooxupé “Brazilian Coffee Scenes Dataset”.

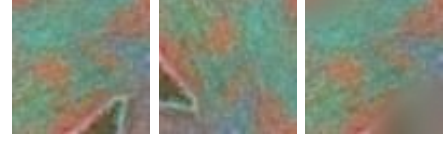


Figura 5: Exemplo de aplicações do efeito das transformações aplicadas nas imagens da base de dados. Na figura a) uma foto de satélite contendo café. b) a figura a) com efeito girar. c) a figura a) com efeito *blur*.

2) Aprendizado Semissupervisionado Baseado em Grafos:

A segunda abordagem considera um conjunto de dados de tamanho N , seleciona uma porcentagem dos dados com rótulo (10%), constrói-se um grafo a partir da semelhança dos exemplos (baseado no algoritmo k NN) e aplica um algoritmo semissupervisionado baseado em grafos (consideramos aqui o LGC - Local and Global Consistency) para prever os rótulos dos outros 90% dos dados.

O algoritmo LGC parte de um dos principais conceitos do aprendizado semissupervisionado, que é a assunção de consistência local e global [11], isto é, pontos próximos (local) e de mesma estrutura (global) são prováveis de possuírem o mesmo rótulo.

Primeiro, dado um grafo ponderado $G(V, E, W)$, onde V representa um conjunto de vértices, E um conjunto de arestas e $W = \{w(e_1), \dots, w(e_k), \dots, w(e_M)\}$ um conjunto de pesos. Sendo w uma função $w : E \rightarrow \mathbb{R}$ que atribui para cada aresta $e_{ij} \in E$ um peso $w(e_{ij})$, ou simplesmente w_{ij} .

Um pseudo-código do algoritmo LGC [11] pode ser conferido na Figura 6 e na Figura 7 há um exemplo de uma rede gerada para propagação de rótulos.

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:

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

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

Figura 6: Pseudo-código do Algoritmo LGC

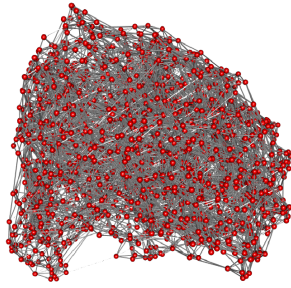


Figura 7: Exemplo do Grafo construído para a propagação.

Assim, ambos os conjuntos de dados gerados são compostos de 10% de dados com rótulos original e 90% de dados com rótulos estimados e serão utilizados para treinar um classificador.

Objetivamos analisar a eficiência e a importância de técnicas de DA na automatização de rotulação de dados. Pois, por exemplo, na aplicação de imagens de satélite da base que utilizamos, cada elemento foi rotulado manualmente por pesquisadores agrícolas, o que é uma tarefa demorada e trabalhosa e além disso, algumas imagens de classes opostas apresentam muita semelhança, dificultando muito a rotulação manual. Com essa proposta espera-se uma boa aproximação com os resultados obtidos no trabalho relacionado que utiliza o mesmo domínio [3]. E algo parecido, mas não superior, com os trabalhos relacionados que realizaram *Data Augmentation*, mas trabalharam com aplicações mais fáceis no sentido de diferenciação entre classes.

V. ANÁLISE EXPERIMENTAL

Para os experimentos iniciais foram geradas bases de dados aumentadas diferentes para cada técnica de DA considerada nesse trabalho. Cada base passou por uma sequência de experimentos de classificação e os resultados são explicitados pelas Figuras 8 e 9. Cada uma dessas Base foi testada com os classificadores *Naive Bayes*, *Multilayer Perceptron*, *Support Vector Machine*, *kNN* e o *J48* (Árvores de Decisão).

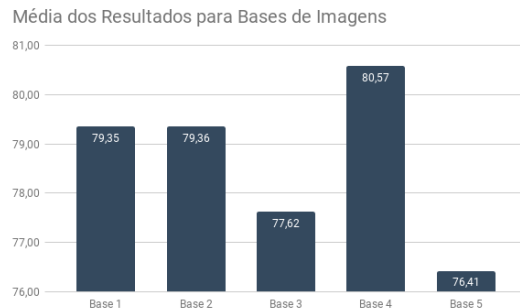


Figura 8: Gráficos do resultado médio de cada Base de Imagens para 5 classificadores diferentes.

Assim, foi escolhida a base com o maior desempenho médio (de cada técnica) para os próximos experimentos de *Deep Learning*. Alguns detalhes dessas bases podem ser conferidos na Tabela I a seguir.

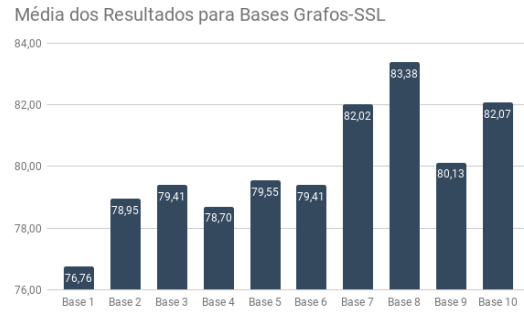


Figura 9: Gráficos dos resultados médio de cada Base de Grafos para 5 classificadores diferentes.

Tabela I: Descrição das bases utilizadas.

Base	Instâncias	Variáveis	Classes
Original	2876	128	2
DA Imagens	2876	128	2
DA Grafos	2876	128	2

Além disso, a Base Original passou por experimentos de *Cross-Validation* com 10-fold, para os 5 classificadores, resultando em uma acurácia média de 84.37% e um desvio padrão de 2.65.

A. Configuração do Algoritmo e do Ambiente Computacional

Os experimentos de classificação foram feitos da seguinte forma:

- 1) Bases Aumentadas 1 e 2 são usadas como Treino para as RN.
- 2) Base Original é usada como Teste para as RN.
- 3) Calcula-se a acurácia.

As bases DA foram testadas com 4 tipos de RN. Começando com os parâmetros padrão de cada rede, variou-se o *Learning Rate*, *Training Time* e o Número de Neurônios em cada camada oculta.

Os parâmetros padrões são *Learning Rate* = 0.3, *Training Time* = 500, *Batch Size* = 100, Número de Neurônios em cada camada oculta = 64.

Para os testes, um parâmetro é variado, enquanto os outros são fixados com os valores padrões. Além disso, cada experimento é realizado diversas vezes e então, é calculada a média da execuções.

As RN empregadas são detalhadas a seguir:

- Rede 1: WEKA com 3 Camadas Ocultas.
- Rede 2: TensorFlow com 3 Camadas Ocultas.
- Rede 3: TensorFlow com 4 Camadas Ocultas (com *Dropout*).
- Rede 4: TensorFlow com 5 Camadas Ocultas.

Os experimentos realizados no WEKA foram executados no Windows 10, com o WEKA 3.8.1. Os experimentos do TensorFlow foram executados no Ubuntu 17.0 com as versões TensorFlow 1.8 e Python 3.6.3. Ambos em um Notebook Samsung Intel I5 7ª geração 2.5GHz, 1 TB de HD, 8 GB de RAM com GPU NVIDIA GEFORCE 920MX.

B. Transformação de Imagens

Para a Base 1 aumentada a partir de Transformação de Imagens, a Tabela II demonstra os resultados para a variação do *Learning Rate*, a Tabela III para o *Training Time* e a Tabela IV para o número de neurônios em cada camada oculta.

Nota-se que o melhor resultado foi 84.46%, na rede de 3 camadas do WEKA, com conjunto de parâmetros: *Learning Rate* = 0.3, *Training Time* = 40 e 64 Neurônios em cada camada oculta.

Tabela II: Variação do *Learning Rate*.

Valor	Rede 1	Rede 2	Rede 3	Rede 4
0.03	83.71	80.38	83.55	81.09
0.05	83.72	80.31	83.81	80.61
0.1	83.90	80.35	83.67	80.49
0.2	83.73	80.05	83.86	80.44
0.3	84.28	79.74	83.97	80.38

Tabela III: Variação do *Epochs/Training Time*.

Valor	Rede 1	Rede 2	Rede 3	Rede 4
40	84.46	83.12	84.02	73.60
100	84.32	82.03	84.14	82.51
200	84.28	81.06	84.05	81.45
400	84.30	80.29	83.46	80.67
500	84.27	79.74	83.97	80.38
600	84.25	79.01	84.29	80.00

Tabela IV: Variação do Números de Neurônios em cada camada oculta.

Valor	Rede 1	Rede 2	Rede 3	Rede 4
32	84.27	80.16	84.18	80.49
64	84.28	79.74	83.80	80.27
128	83.94	80.06	83.80	80.27
256	84.11	80.24	83.98	80.16
512	84.01	80.15	83.80	80.04

C. Aprendizado Semissupervisionado Baseado em Grafos

Para a Base 2 aumentada a partir da abordagem de Aprendizado Semissupervisionado Baseado em Grafos, a Tabela V apresenta os resultados para a variação do *Learning Rate*, a Tabela VI para o *Training Time* e a Tabela VII para o número de neurônios em cada camada.

Nota-se também que o maior resultado foi 84.11% com a rede de 3 camadas ocultas do WEKA, com *Learning Rate* = 0.3, *Training Time* = 600 e 64 Neurônios em cada camada oculta.

Tabela V: Variação do *Learning Rate*.

Valor	Rede 1	Rede 2	Rede 3	Rede 4
0.03	81.47	82.89	83.83	82.98
0.05	83.55	82.88	83.77	82.86
0.1	83.73	82.72	83.80	82.99
0.2	83.38	82.85	83.77	82.68
0.3	82.61	82.68	83.70	82.80

Tabela VI: Variação do *Epochs/Training Time*.

Valor	Rede 1	Rede 2	Rede 3	Rede 4
40	82.78	82.30	83.33	82.25
100	82.79	82.62	83.92	82.78
200	82.37	82.94	83.74	82.94
400	83.38	83.07	83.75	82.98
500	82.64	83.09	83.81	83.14
600	84.11	82.97	83.78	83.18

Tabela VII: Variação do Números de Neurônios em cada camada oculta.

Valor	Rede 1	Rede 2	Rede 3	Rede 4
32	83.31	83.00	83.71	83.08
64	82.61	83.09	83.82	83.14
128	82.79	82.83	83.81	83.03
256	81.43	82.91	83.80	82.89
512	83.76	83.20	83.79	76.49

D. Comparação das Técnicas

Feito os experimentos de variação de parâmetros, é questionado se alguma técnica de DA se saiu melhor que a outra para a tarefa de classificação. Sendo assim, novos experimentos foram realizados.

Para essa nova leva de experimentos, os melhores parâmetros de *Learning Rate* e Número de Neurônios em cada camada foram considerados para as duas técnicas. Além disso, escolheu-se a rede com mais camadas ocultas, a Rede 4 (com 5 camadas ocultas).

Dessa forma, a Base 1 aumentada por Transformação de Imagens e a Base 2 aumentada por Aprendizado Semissupervisionado passam por uma sequência de 30 repetições de experimentos com a Rede 4 com *Learning Rate* = 0.3, 64 neurônios em cada camada oculta e *Training Time* = 500, dentre outros parâmetros fixados, com as sementes sendo definidas pelo tempo.

Deste modo, gerou-se uma amostra das acurácias obtidas de tamanho $N = 30$ para cada base. Um resumo dos resultados obtidos pode ser conferido na Tabela VIII.

Tabela VIII: Resultados obtidos para as duas bases com a Rede 4.

Medida	Base DA Imagens	Base DA Grafos
Média	80.30	82.91
Mediana	80.26	82.94
Desvio Padrão	0.2769	0.2758
Variância	0.0782	0.0741

Fazendo um teste de hipótese do tipo *Paired t-test* para essa amostra de experimentos, temos:

- Médias das amostras $\bar{X}_1 = 80.30$, $\bar{X}_2 = 82.91$;
- Desvio padrão das amostras: $s_1 = 0.2769$ e $s_2 = 0.2758$;
- Tamanho das amostras $N = 30$;
- $\bar{D} = -2.609$ e $s_D = 0.421$, onde \bar{D} é a diferença das médias ($\bar{D} = \bar{X}_1 - \bar{X}_2$) e s_D é a diferença dos desvios ($\bar{D} = X_1^i - X_2^i$, para $i = 1, 2, \dots, N$).

Vamos definir:

$$H_0 : \mu_D = 0 \quad (1)$$

$$H_a : \mu_D \neq 0 \quad (2)$$

Isto é, um teste bicaudal, para verificar se há diferença significativa na média das duas técnicas, onde H_0 é a hipótese nula de que as médias são iguais e H_a , a hipótese alternativa, de que há uma diferença nas médias. Além disso, o nível de significância é definido como $\alpha = 0.05$.

Então, temos que o valor crítico para o teste é $t_c = 2.045$ e a região de rejeição é $R = \{t : |t| > 2.045\}$. Dessa forma:

$$t = \frac{\bar{D}}{s_D/\sqrt{N}} = \frac{-2.609}{0.421/\sqrt{30}} = -33.973 \quad (3)$$

Como $|t| = 33.973 > t_c = 2.045$ a hipótese nula é rejeitada. Também, utilizando o p -valor, temos $p = 0 < 0.05$. Sendo assim, podemos concluir que há uma diferença significativa entre as médias das duas amostras com um intervalo de confiança de 95%.

E. Discussão

Fazendo uma análise dos resultados obtidos nota-se que a Rede do WEKA obteve, em geral, melhores resultados em relação as Redes do TensorFlow com 3 ou mais camadas. Isso pode ser explicado pelo fato de outros parâmetros mais específicos não terem sido considerados e setados como iguais para todas as redes.

Analisando as redes do TensorFlow (Redes 2, 3 e 4) pode-se concluir que a Rede 3 (Rede com 4 Camadas Ocultas com *Dropout*) obteve melhores resultados em relação a Rede 2 e Rede 4. Sendo assim, a técnica de *dropout* pode ser considerada quando se trabalha com Redes Neurais Profundas.

Em relação as duas técnicas de DA, ambas se saíram eficientes para todos os testes realizados, chegando a 84.46% de acurácia para a técnica baseada em Transformações de Imagens e 84.11% para a técnica baseada em Aprendizado Semissupervisionado. Enquanto que com uma Rede Neural padrão com 1 camada oculta, os resultados obtidos anteriormente tinham sido 82.58% e 83.17%, respectivamente.

VI. CONCLUSÕES

Pode-se concluir que ambas as técnicas de DA aqui consideradas são eficientes para a aplicação escolhida, sendo assim, técnicas de DA podem ser aplicadas para a automatização da rotulagem de dados e para melhorar o modelo gerado pelos algoritmos supervisionado, resultando numa melhor acurácia. Ficando a cargo da aplicação ou preferência a escolha de qual técnica utilizar.

Dentre as vantagens da técnica de Transformação de Imagens é a possibilidade de utilização em qualquer aplicação que tenha imagens. Como desvantagem, deve-se tomar cuidado com a redundância (caso o efeito aplicado seja mínimo e as imagens fiquem praticamente iguais, pode não resultar em uma melhora na acurácia) ou na perda de sentido (caso o efeito aplicado seja excessivo e a imagem não pareça mais com a classe desejada).

Para a abordagem de Aprendizado Semissupervisionado, as vantagens são devido ao fato de ser uma abordagem ampla, pois pode-se aplicar diversos algoritmos de propagação de

rótulos ou alterações na construção do grafo. Como desvantagem, só pode ser utilizado em aplicações com elementos rotulados e vários elementos não rotulados. Também, um testes de hipótese revelou uma diferença significativa na média das duas melhores bases de cada técnica, em favor para a abordagem de Grafos-SSL.

Além disso, o aprendizado profundo com adição de técnicas como regularizadores poderiam ser utilizadas para melhorar os resultados obtidos.

Para trabalhos futuros, focando na abordagem de Aprendizado Semissupervisionado baseado em Grafos, iremos fazer o estudo de diferentes algoritmos de propagação de rótulos, fazendo um comparativo entre eles.

REFERÊNCIAS

- [1] Olivier Chapelle, Bernhard Schölkopf, and Alexander Zien. *Semi-Supervised Learning*. The MIT Press, 1st edition, 2010.
- [2] Ian Goodfellow, Yoshua Bengio, Aaron Courville, and Yoshua Bengio. *Deep learning*, volume 1. MIT press Cambridge, 2016.
- [3] Otávio AB Penatti, Keiller Nogueira, and Jefersson A dos Santos. Do deep features generalize from everyday objects to remote sensing and aerial scenes domains? In *Computer Vision and Pattern Recognition Workshops (CVPRW)*, 2015 *IEEE Conference on*, pages 44–51. IEEE, 2015.
- [4] Maria Carolina Monard and José Augusto Baranauskas. Conceitos sobre aprendizado de máquina. *Sistemas inteligentes-Fundamentos e aplicações*, 1(1):32, 2003.
- [5] Simon Haykin. *Redes neurais: princípios e prática*. Bookman Editora, 2007.
- [6] Nitish Srivastava, Geoffrey Hinton, Alex Krizhevsky, Ilya Sutskever, and Ruslan Salakhutdinov. Dropout: A simple way to prevent neural networks from overfitting. *The Journal of Machine Learning Research*, 15(1):1929–1958, 2014.
- [7] Jia Shijie, Wang Ping, Jia Peiyi, and Hu Siping. Research on data augmentation for image classification based on convolution neural networks. In *Chinese Automation Congress (CAC)*, 2017, pages 4165–4170. IEEE, 2017.
- [8] Jason Wang and Luis Perez. The effectiveness of data augmentation in image classification using deep learning. Technical report, Technical report, 2017.
- [9] Emmanuel Okafor, Rik Smit, Lambert Schomaker, and Marco Wiering. Operational data augmentation in classifying single aerial images of animals. In *INnovations in Intelligent Systems and Applications (INISTA)*, 2017 *IEEE International Conference on*, pages 354–360. IEEE, 2017.
- [10] Olga Zoidi, Anastasios Tefas, Nikos Nikolaidis, and Ioannis Pitas. Iterative label propagation on facial images. In *Signal Processing Conference (EUSIPCO)*, 2014 *Proceedings of the 22nd European*, pages 1222–1226. IEEE, 2014.
- [11] Denny Zhou, Olivier Bousquet, Thomas N Lal, Jason Weston, and Bernhard Schölkopf. Learning with local and global consistency. In *Advances in neural information processing systems*, pages 321–328, 2004.