XXVIII Congresso {virtual} de Iniciação Científica da Unicamp 2020

Classificação de imagens biológicas usando descritores extraídos por redes convolucionais profundas

Marina Rocha*, João Batista Florindo

RESUMO

Redes neurais convolucionais têm sido muito relevantes na área de classificação de imagens. Neste projeto, elas foram utilizadas na categorização de imagens biológicas, por meio da extração de descritores e votação entre diferentes classificadores (*ensemble*), utilizando estratégias de combinação de descritores de baixo e alto nível para aumentar a acurácia. A metodologia desenvolvida foi testada na base "KTH-TIPS-2b" (9), confirmando sua eficiência diante de problemas típicos de classificação de imagens, e então ela foi aplicada a dois problemas práticos de grande importância. O primeiro é a identificação de espécies de plantas brasileiras a partir de imagens escaneadas da superfície foliar (2). O segundo problema é a identificação e categorização de três tipos de cistos bucais (3). O *ensemble* resultou em um aumento na taxa de acerto e a estratégia de combinação de descritores se mostrou promissora, resultando em um sistema computacional capaz de auxiliar no trabalho de médicos e botânicos na categorização de cistos bucais e espécies de plantas, respectivamente.

Palavras-chave: Rede convolucional profunda, ensemble de classificadores, classificação de imagens biológicas.

1. Contexto e objetivos

As redes neurais são uma poderosa ferramenta computacional, que permeia praticamente todos os aspectos da sociedade moderna. Atualmente, o interesse nelas vem crescendo, principalmente nas redes convolucionais profundas. Por mostrar desempenho superior a outros modelos, têm sido aplicadas em diversas áreas, como classificação de imagens médicas (8) e predição de sequências de DNA-RNA (1), entre muitas outras.

Uma característica importante das redes neurais é que elas são fortemente hierarquizadas e flexíveis, sendo capazes de extrair descritores mesmo de imagens complexas (5). Este projeto utilizou essa propriedade na extração de descritores em redes convolucionais profundas e os combinou com diversos classificadores.

A metodologia desenvolvida foi aplicada à análise de imagens biológicas, uma das áreas mais proeminentes em fornecer dados complexos e em larga escala. Por isso, são apropriadas para a análise por redes profundas, propícias para detectar e discriminar padrões escondidos que não são identificáveis por descritores clássicos e muito menos pelo olho humano.

1.1. Categorização de texturas sob condições distintas de iluminação, posição e escala

Um dos problemas em visão computacional é a classificação de imagens independentemente das condições sob as quais as fotos foram tiradas. Fatores externos podem alterar drasticamente uma fotografia. No entanto, espera-se que a rede seja capaz de classificar a imagem corretamente, mesmo submetida a essas alterações.

Para analisar esse aspecto utilizou-se a base de dados conhecida como "KTH-TIPS-2b" (9), que consiste em um conjunto de fotos de texturas em que cada classe é dividida em quatro subclasses, submetidas a variações na iluminação, posição e escala.

A metodologia utilizada consiste em escolher uma dessas amostras de cada classe e usar apenas esse conjunto para treinar a rede, enquanto as outras três subclasses são usadas no teste. Com isso, é possível analisar qual é o resultado da rede ao se deparar com conjuntos de teste submetidos a condições ainda não vistas.

1.2. Identificação de espécies de plantas

Quando feita por especialistas, a identificação de espécies de plantas costuma focar na análise das flores e frutos, por estes serem órgãos de fácil observação. No entanto, essas estruturas nem sempre estão presentes. Por isso é mais conveniente analisar as superfícies foliares, devido à sua disponibilidade em qualquer época. Utilizamos a base 1200Tex proposta e estudada em (2).

O problema é que as folhas são muito heterogêneas, havendo variação significativa inclusive quando uma mesma planta é submetida a condições ambientais distintas. Este cenário exige um ferramental computacional avançado e as redes profundas têm demonstrado capacidade para esta tarefa.

1.3. Categorização de cistos bucais

Cistos odontogênicos são lesões da mandíbula que podem ser divididas em três grupos: radiculares, queratocistos esporádicos e queratocistos sindrômicos. A identificação do tipo de cisto é fundamental para fins de diagnóstico médico e tratamento.

Embora atributos que diferenciem tais cistos sejam bem conhecidos por histopatologistas, o processo neste caso é manual, fortemente sujeito à subjetividade do diagnóstico humano, além

de não permitir análise de um grande conjunto de dados em um intervalo de tempo razoável.

Para preencher esta lacuna, métodos computacionais vêm sendo aplicados nos últimos anos, alcançando resultados promissores, em particular com as redes neurais convolucionais profundas (8).

Fig. 1. Exemplos de imagens das bases de texturas, folhas e cistos, respectivamente.

2. Metodologia

2.1. Redes convolucionais

Uma rede neural é um método de classificação que se baseia em reconhecer padrões de bases de treino para categorizar novos dados de forma rápida, automatizada e eficiente.

Em particular, as redes neurais convolucionais foram desenvolvidas para processar dados na forma de matrizes, especialmente imagens. Elas são formadas por camadas convolucionais e camadas de condensamento intercaladas, como no esquema da Figura 3. Este tipo de rede apresenta algumas características principais:

- **Conexões locais (*feature maps*):** analisamos características locais das imagens, mantendo a estrutura original da matriz de pixels.
- **Pesos compartilhados:** cada *feature map* possui um conjunto fixado de pesos, como um filtro deslocado ao longo de toda a matriz original. O intuito dessa propriedade é de que um filtro otimizado para identificar algo, possa fazê-lo em qualquer lugar da imagem.
- **Condensamento (*pooling*):** mescla características similares, reduzindo a dimensão da matriz.
- **Múltiplas camadas (*deep learning*):** redes profundas exploram a propriedade de que muitas estruturas naturais possuem uma hierarquia intrínseca, ou seja, características de alto nível de escala (globais) estão também presentes em níveis mais baixos (locais).

A camada convolucional é embasada no operador de convolução, adaptado para o caso discreto bidimensional (5):

$$S(i,j) = (K * I)(i,j) = \sum_m \sum_n I(i-m, j-n)K(m,n),$$

em que K representa o *kernel* da rede, I corresponde ao dado que está sendo fornecido como entrada e S às conexões locais.

Após a camada de convolução, é aplicada uma função de ativação (ReLU), que mapeia os valores recebidos para um novo

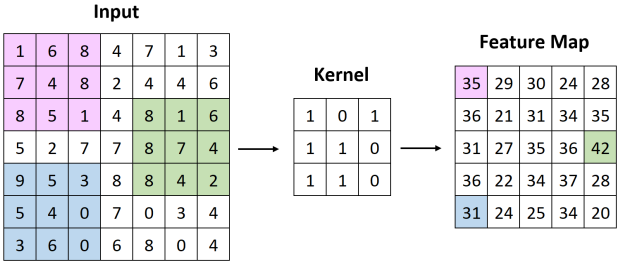

Fig. 2. Esquema de uma aplicação convolucional.

intervalo. Por fim, a camada de condensamento aplica o critério de *max pooling*, que escolhe o maior número de uma vizinhança para representá-la..

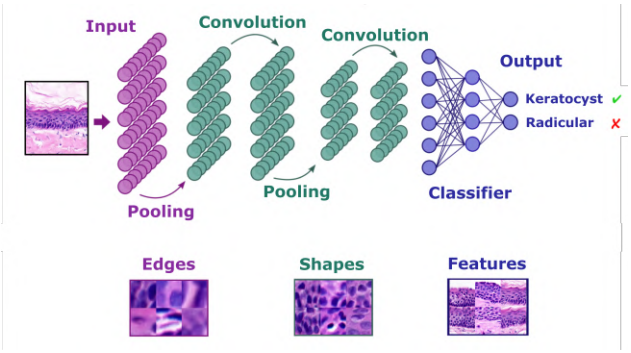

Fig. 3. Esquema genérico de uma rede neural convolucional.

O caráter hierárquico dessas redes faz com que as camadas iniciais extraiam as características mais básicas da imagem (7), tornando possível utilizar estratégias como a de *transfer learning*, que consiste em aproveitar pesos otimizados em bases de dados externas significativamente maiores (ImageNet).

2.2. Classificadores

As últimas camadas de uma rede convolucional também são conhecidas como "classificadores". Elas recebem os descritores das imagens e determinam uma classificação para cada imagem. Neste projeto foram utilizados os seguintes classificadores:

- ***Fully Connected Layer (FCL):*** fileira simples de neurônios totalmente conectados.
- ***Support Vector Machine (SVM):*** tenta encontrar um hiperplano em um espaço N-dimensional que melhor separe os pontos que representam os dados (11).
- ***K-Nearest Neighbors (KNN):*** verifica a classe dos K descritores mais próximos à imagem a ser classificada e categoriza de acordo com a classe mais recorrente entre estes K vizinhos.
- ***Random Forest (RF)*:** constroi múltiplas árvores de decisão e as combina para obter uma classificação mais acurada.
- ***Linear Discriminant Analysis (LDA):*** procura uma combinação linear de variáveis que melhor separa duas classes

(12). É um método estatístico que usa parâmetros como expectativa e covariância.

Além desses, outros três classificadores também foram implementados. Eles usam uma estratégia que consiste em combinar descritores de baixo nível com outros de alto nível (4), beaseando-se na premissa de que a combinação desses descritores resultaria no aumento da acurácia. Foi escolhido o método LBP (*Local Binary Pattern*) como extrator de características de baixo nível, já que a rede convolucional pode ser considerada como um método de alto nível.

Temos os vetores $\overrightarrow{d}$ (proveniente de uma relação de distâncias entre imagens a partir do método LBP) e $\overrightarrow{r}$ (confiança do classificador SVM), que representam descritores de nível baixo e alto, respectivamente. Esses vetores foram combinados usando três métodos diferentes, cada um deles resultando em um novo classificador.

– **SVM + LBP:** combinação usando as distâncias para calcular pesos (w_i) para o vetor de confiança ($\overrightarrow{r}$):

$$w_i = \left(\frac{1}{m} \sum_{\substack{j=1...m \\ j \neq i}} d_j \right) / d_i$$

$$\overrightarrow{r_w} = [w_1 r_1, w_2 r_2, \ldots, w_m r_m]$$

– **SVM + LBP + NN:** o vetor de confiança ($\overrightarrow{r}$) foi concatenado com o vetor de distâncias ($\overrightarrow{d}$) e usado como entrada para treinar uma rede do tipo FCL, que determinou a melhor combinação entre os valores de confianças e distâncias.

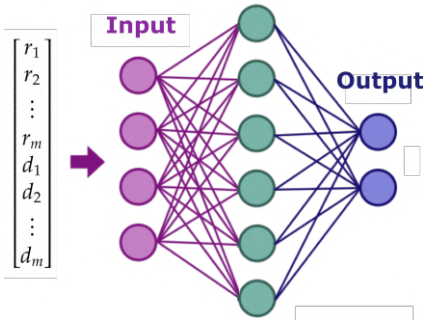

Fig. 4. Visualização do método SVM + LBP + NN.

– **SVM + LDA + LBP + NN:** mesmo princípio do classificador anterior, mas concatena os vetores de confiança dos métodos SVM e LDA com o vetor de distâncias.

2.3. Comitê de máquinas (Ensemble):

Cada um desses oito classificadores determinou a qual classe cada imagem de teste pertence. A partir disso, foi feito um sistema de votação, conhecido como comitê de máquinas (10), em que a escolha de cada classificador representa um voto. O intuito foi aumentar a acurácia final, escolhendo a classe que mais apareceu entre os resultados dos oito classificadores.

Essa estratégia é promissora, já que ela não demanda uma grande base de dados para treino nem um alto poder computacional. Essas características são essenciais para as bases estudadas neste projeto, considerando que essas imagens são de difícil obtenção. Além disso, tanto o uso da combinação de descritores quanto o do comitê de máquinas são muito convenientes, pois possibilitam melhorar os resultados de acurácia sem a necessidade de incluir novas informações externas ou de buscar por outras abordagens completamente diferentes.

3. Resultados e Análises

Todas as redes foram implementadas em *Python*. Foram utilizados os modelos pré-treinados, ResNet (6) e AlexNet. As abordagens para os parâmetros da rede foram:

– **Ajuste fino (*finetuning*):** depois da rede ser pré-treinada na base externa ImageNet, todos os parâmetros são otimizados usando as bases de dados em questão neste projeto.
– **Extração fixa de características:** a rede neural também é pré-treinada na ImageNet, porém as primeiras camadas mantêm os pesos "congelados", otimizando apenas os pesos das camadas finais nas bases internas.

Os conjuntos de teste e treino foram separados aleatoriamente, numa proporção de metade das fotos para cada grupo. Os experimentos foram repetidos dez vezes, alterando aleatoriamente as imagens de treino, e então foram calculados os desvios padrão e as acurácias máximas médias.

3.1. Categorização de texturas

Quando as subclasses da base são ignoradas, ou seja, todos os tipos de amostras estão presentes no conjunto de treino, a rede chegou a praticamente 100% de acerto. Já com cada um dos tipos de amostras como conjunto de treino e os demais como teste, o resultado máximo obtido ficou em torno de 75%. Isso já era previsto, pois no segundo caso a rede já tinha tido contato com imagens similares durante o treino.

Esses resultados podem ser considerados satisfatórios, já que eles representam o desempenho da rede diante de imagens submetidas a condições adversas. Isso significa que a metodologia implementada foi capaz de fazer uma boa generalização, mantendo uma taxa de acerto alta mesmo diante de imagens que possuem características não presentes na base de treino, como variações de iluminação, posição ou escala.

3.2. Identificação de espécies de plantas

Com a ResNet, a maior taxa de acerto foi obtida pelo *ensemble*, enquanto que com a AlexNet os melhores resultados foram obtidos pelo classificador SVM + LBP + NN. Comparando as duas estruturas, a ResNet teve um desempenho consideravelmente melhor.

Table 1. Superfícies foliares: acurácia máxima média com seu respectivo desvio padrão.

(%)	ResNet		AlexNet	
Classificador	*Finetuning*	Fixa	*Finetuning*	Fixa
FCL	87 ± 1	75 ± 1	83 ± 1	82 ± 1
LDA	79 ± 2	47 ± 2	70 ± 2	61 ± 2
SVM	88 ± 1	80 ± 1	73 ± 2	69 ± 2
RF	76 ± 2	56 ± 2	56 ± 4	52 ± 1
KNN	84 ± 1	75 ± 2	67 ± 1	60 ± 1
SVM + LBP	89 ± 1	77 ± 2	73 ± 1	81 ± 1
SVM + LBP + NN	90 ± 1	81 ± 2	**84 ± 2**	**83 ± 2**
SVM + LDA + LBP + NN	87 ± 2	79 ± 1	78 ± 2	73 ± 1
ENSEMBLE	**91 ± 1**	**83 ± 1**	80 ± 1	73 ± 2

O desempenho geral dos três classificadores com combinação de descritores foi consistentemente superior ao dos demais. Isso pode ser um indício para a corroboração da premissa de que essa combinação aumenta a acurácia.

Outra forma de visualizar os resultados é através da matriz de confusão, que mostra graficamente o número de imagens atribuídas a cada classe. Na Figura 5 temos a comparação entre as matrizes de confusão utilizando os métodos de ajuste fino e de extração fixa de características. Ficou evidente o resultado superior do primeiro, o que era esperado, considerando-se que o tipo de imagem analisada é muito específico e, portanto, a base de dados externa não teria tanto a acrescentar no desempenho da rede.

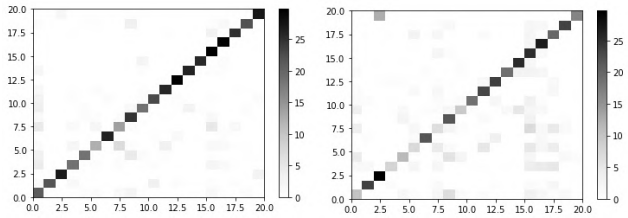

Fig. 5. Representação em escala de cinza das matrizes de confusão.

3.3. Categorização de cistos bucais

Para esta base, além das variações já utilizadas anteriormente, foram feitos alguns outros experimentos relevantes. Considere **k** para a classe "esporádicos", **s** para "sindrômicos" e **r** para "radiculares". Portanto, além do teste padrão entre as três classes (k x s x r), testamos a capacidade da rede de distinguir entre os dois tipos de queratocistos (k x s) e entre os radiculares e os queratocistos (ks x r).

Novamente, a ResNet teve um desempenho superior, assim como a estratégia de ajuste fino. Os resultados foram condizentes com o esperado, já que a diferença entre radiculares e queratocistos (ks x r) é a mais acentuada, sendo portanto a classificação mais fácil, chegando a quase 100% de acerto quando utilizada a ResNet com ajuste fino.

References

[1] Babak Alipanahi, Andrew Delong, Matthew T. Weirauch, and Brendan J. Frey. Predicting the sequence specificities of DNA- and RNA-binding proteins by deep learning. *NATURE BIOTECHNOLOGY*, 33(8):831+, AUG 2015.

[2] Dalcimar Casanova, Jarbas Joaci de Mesquita Sá Junior, and Odemir Martinez Bruno. Plant leaf identification using gabor wavelets. *International Journal of Imaging Systems and Technology*, 19(3):236–243, 2009.

[3] Joao B. Florindo, Odemir M. Bruno, and Gabriel Landini. Morphological classification of odontogenic keratocysts using bouligand-minkowski fractal descriptors. *Computers in Biology and Medicine*, 81:1 – 10, 2017.

[4] JI Forcén, M Pagola, E Barrenechea, and H Bustince. Combination of features through weighted ensembles for image classification. *Applied Soft Computing*, 84:105698, 2019.

[5] Ian Goodfellow, Yoshua Bengio, and Aaron Courville. *Deep Learning*. MIT Press, 2016. http://www.deeplearningbook.org.

[6] Kaiming He, Xiangyu Zhang, Shaoqing Ren, and Jian Sun. Deep residual learning for image recognition. In *Proceedings of the IEEE conference on computer vision and pattern recognition*, pages 770–778, 2016.

[7] Yann LeCun, Yoshua Bengio, and Geoffrey Hinton. Deep learning. *nature*, 521(7553):436, 2015.

[8] Geert Litjens, Thijs Kooi, Babak Ehteshami Bejnordi, Arnaud Arindra Adiyoso Setio, Francesco Ciompi, Mohsen Ghafoorian, Jeroen A. W. M. van der Laak, Bram van Ginneken, and Clara I. Sanchez. A survey on deep learning in medical image analysis. *MEDICAL IMAGE ANALYSIS*, 42:60–88, DEC 2017.

[9] P Mallikarjuna, Alireza Tavakoli Targhi, Mario Fritz, Eric Hayman, Barbara Caputo, and Jan-Olof Eklundh. The kth-tips2 database. *Computational Vision and Active Perception Laboratory (CVAP), Stockholm, Sweden*, 2006.

[10] Lior Rokach. Ensemble-based classifiers. *Artificial Intelligence Review*, 33(1-2):1–39, 2010.

[11] Bernhard Scholkopf and Alexander J Smola. *Learning with kernels: support vector machines, regularization, optimization, and beyond*. MIT press, 2001.

[12] Jieping Ye, Ravi Janardan, and Qi Li. Two-dimensional linear discriminant analysis. In *Advances in neural information processing systems*, pages 1569–1576, 2005.

Table 2. Cistos: acurácia obtida usando a estrutura ResNet.

(%)	k x s		k x s x r		ks x r	
Classificador	*Finetuning*	Fixa	*Finetuning*	Fixa	*Finetuning*	Fixa
FCL	82 ± 4	77 ± 4	**86 ± 3**	76 ± 2	**99 ± 1**	93 ± 3
LDA	77 ± 7	74 ± 3	80 ± 4	70 ± 3	96 ± 2	90 ± 2
SVM	80 ± 6	73 ± 7	83 ± 4	76 ± 3	97 ± 2	93 ± 3
RF	77 ± 5	69 ± 3	81 ± 4	65 ± 3	96 ± 3	76 ± 3
KNN	77 ± 9	73 ± 5	81 ± 4	74 ± 4	93 ± 4	91 ± 3
SVM + LBP	81 ± 7	74 ± 6	83 ± 4	76 ± 5	96 ± 2	93 ± 3
SVM + LBP + NN	82 ± 5	77 ± 4	84 ± 3	76 ± 4	**99 ± 1**	94 ± 3
SVM + LDA + LBP + NN	80 ± 5	77 ± 4	83 ± 4	77 ± 4	98 ± 1	94 ± 3
ENSEMBLE	**83 ± 6**	**80 ± 4**	**86 ± 3**	**80 ± 3**	**99 ± 1**	**95 ± 3**

Table 3. Cistos: acurácia obtida usando a estrutura AlexNet.

(%)	k x s		k x s x r		ks x r	
Classificador	*Finetuning*	Fixa	*Finetuning*	Fixa	*Finetuning*	Fixa
FCL	71 ± 7	**77 ± 4**	**77 ± 6**	**79 ± 3**	**88 ± 9**	**95 ± 1**
LDA	67 ± 6	65 ± 7	51 ± 7	60 ± 5	76 ± 7	80 ± 3
SVM	68 ± 7	65 ± 6	61 ± 6	66 ± 3	81 ± 6	87 ± 3
RF	67 ± 6	68 ± 4	63 ± 8	64 ± 4	81 ± 5	80 ± 3
KNN	59 ± 9	68 ± 4	59 ± 7	65 ± 4	79 ± 5	83 ± 4
SVM + LBP	72 ± 7	67 ± 5	64 ± 7	66 ± 5	81 ± 7	86 ± 3
SVM + LBP + NN	69 ± 9	68 ± 5	62 ± 6	67 ± 3	81 ± 7	88 ± 4
SVM + LDA + LBP + NN	70 ± 6	70 ± 6	59 ± 7	67 ± 4	81 ± 6	85 ± 2
ENSEMBLE	**74 ± 8**	75 ± 2	69 ± 7	74 ± 3	84 ± 7	90 ± 3