



UNIVERSIDADE ESTADUAL DE CAMPINAS
INSTITUTO DE MATEMÁTICA, ESTATÍSTICA E COMPUTAÇÃO CIENTÍFICA
DEPARTAMENTO DE MATEMÁTICA APLICADA



MARINA ROCHA

Classificação de imagens biológicas usando descritores extraídos por redes convolucionais profundas*

Monografia apresentada ao Instituto de Matemática, Estatística e Computação Científica da Universidade Estadual de Campinas como parte dos requisitos para obtenção de créditos na disciplina Projeto Supervisionado, sob a orientação do(a) Prof. Dr. João Batista Florindo.

Campinas
2019

*Este trabalho foi financiado pelo CNPq, PIBIC/2019.

Resumo

Com a crescente relevância das Redes Neurais Profundas para resolver o problema de classificação de imagens, diversas áreas de aplicação têm surgido, uma delas sendo a categorização de imagens biológicas. O presente projeto aplicou estes conceitos na área médica, na discriminação de cistos bucais, na botânica, para a identificação de espécies de plantas a partir de imagens da superfície foliar e em texturas sob condições diversas de iluminação, ângulo e posição. O projeto consiste em utilizar redes neurais profundas, em particular as redes convolucionais, para extrair os descritores de imagens, então combiná-los com diferentes classificadores, por meio de um sistema de votação (*ensemble*) para decidir a classe da imagem. A extração de características se mostrou muito eficiente, enquanto que a estratégia de *ensembles* acrescentou pouco à acurácia das redes.

Rede convolucional profunda; *Ensemble* de classificadores; Classificação de imagens biológicas.

1 Introdução

Um dos tópicos que mais está em destaque nos dias de hoje é o conceito de “rede neural”, principalmente as redes convolucionais, que têm se mostrado muito eficientes, por exemplo, para problemas de classificação. Esse método tem sido aplicado nas mais diversas áreas, como filtros de conteúdos em redes sociais [4], em imagens médicas em geral [6] e na predição de sequências de DNA-RNA [1].

Uma característica importante das redes neurais é que elas são fortemente hierarquizadas e flexíveis, o que, no caso específico da análise de imagens, as torna capazes de extrair descritores sob diferentes escalas e com uma grande capacidade de representação de objetos representados nestas imagens. O intuito deste projeto é utilizar esta propriedade na extração de descritores de imagens em redes convolucionais profundas e combiná-los com diversos classificadores diferentes (*Fully Connected Layer* - FCL, *Support Vector Machine* - SVM, *K-Nearest Neighbors* - KNN, *Random Forest* - RF e *Linear Discriminant Analysis* - LDA).

A metodologia desenvolvida foi aplicada a três problemas práticos de grande relevância e que envolvem a análise de imagens biológicas. O primeiro é a identificação de espécies de plantas brasileiras a partir de imagens escaneadas da superfície foliar [2]. Este é um problema naturalmente desafiador, uma vez que as estruturas classicamente usadas para este fim são flores e frutos. Ao mesmo tempo, estes órgãos muitas vezes não são facilmente acessíveis na natureza e a possibilidade de se usar a folha é um grande facilitador para este processo.

O segundo problema é a identificação e categorização de três tipos de cistos bucais [3]. Este também é um problema complexo pois a inspeção visual muitas vezes se mostra imprecisa e custosa. Neste sentido, a automatização do processo é muito bem vinda como uma ferramenta que auxilia o diagnóstico médico, além de permitir que o tratamento possa ser antecipado, aumentando assim as chances de uma boa resposta clínica para o paciente.

O terceiro problema é o de reconhecimento de texturas, usando a base “KTH-TIPS2-b” para testar o desempenho da rede profunda diante de variações que não foram vistas durante o treino, como mudanças de ângulo e iluminação.

2 Redes neurais profundas

2.1 Redes neurais artificiais

Redes neurais artificiais são um sistema computacional inspirado em neurônios biológicos. O conceito básico é o de executar uma tarefa específica usando exemplos como base de treino. Funciona como um aprendizado de máquina, em que otimiza-se uma função de custo para minimizar o erro da rede.

Para fazer esse processo de otimização da rede, utiliza-se uma estratégia conhecida como “gradiente descendente”. A cada etapa, os parâmetros (pesos) do sistema são atualizados, até atingir algum determinado critério de parada.

No caso do presente projeto, a tarefa que a rede executa é a de classificação de imagens. Sendo assim, esses pesos são utilizados em uma parte da base de imagens da qual não se conhece a classificação *a priori* (conjunto de teste), com o intuito de que a rede tenha uma boa generalização e também atribua classificações corretas a este novo conjunto.

A estrutura de uma rede neural basicamente consiste em camadas de “neurônios” conectados entre si, como em um grafo. As entradas podem ser os valores RGB de cada pixel, as camadas intermediárias são tipicamente responsáveis pela extração de descritores da imagem, através da otimização dos pesos e da aplicação de funções não lineares (Ex.: sigmoide, ReLU), enquanto que as últimas camadas costumam agir como classificadores [4].

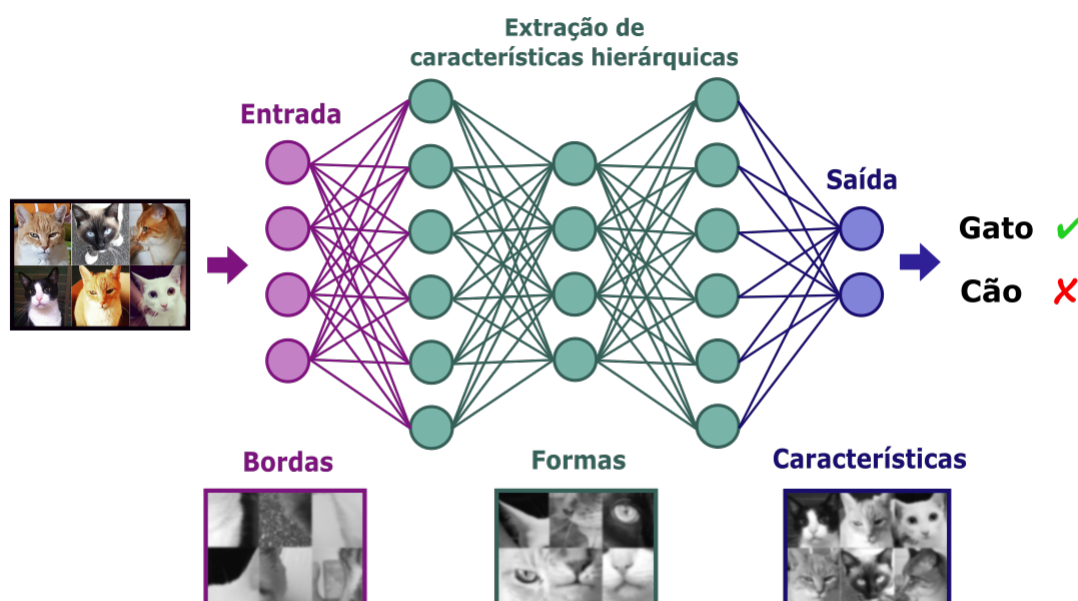


Figura 1: Esquema de uma rede neural simples.

Expandindo o conceito de redes neurais, definem-se as redes profundas, que basicamente são redes neurais artificiais que possuem um grande número de camadas. Neste tipo de rede não é possível rastrear o que o algoritmo está fazendo, por isso o método é visto como uma “caixa preta”. No entanto, é este tipo de rede que tem apresentado os resultados mais satisfatórios em análise de imagens [8].

2.2 Redes neurais convolucionais

As redes convolucionais são formadas por uma estrutura diferenciada, que foi projetada para processar dados na forma de matrizes, especialmente imagens. Este tipo de rede apresenta algumas características principais:

- **Conexões locais (*feature maps*):** nas redes convolucionais, são analisadas características locais das imagens, utilizando submatrizes conhecidas como *feature maps*, mantendo a estrutura original da matriz de pixels.
- **Pesos compartilhados:** cada *feature map* possui um conjunto fixado de pesos, que atuam como um filtro sobre a imagem. Esse mesmo filtro é aplicado ao longo de toda a matriz original, sendo apenas deslocado. A ideia por trás dessa propriedade é a de que um filtro otimizado para identificar rostos, por exemplo, deva ser capaz de identificar um rosto em qualquer lugar da imagem.
- **Condensamento (*pooling*):** o papel da camada de condensamento (*pooling layer*) é mesclar características similares, ou seja, reduzir a dimensão da matriz. Um desses métodos é o *max pooling*, que consiste em escolher o número de maior valor.
- **Múltiplas camadas (*deep learning*):** todo esse processo de montar os *feature maps*, otimizar os pesos compartilhados, aplicar uma função não-linear e condensar a matriz é repetido várias vezes ao longo da rede. Isso caracteriza as múltiplas camadas de aprendizagem (*deep learning*).

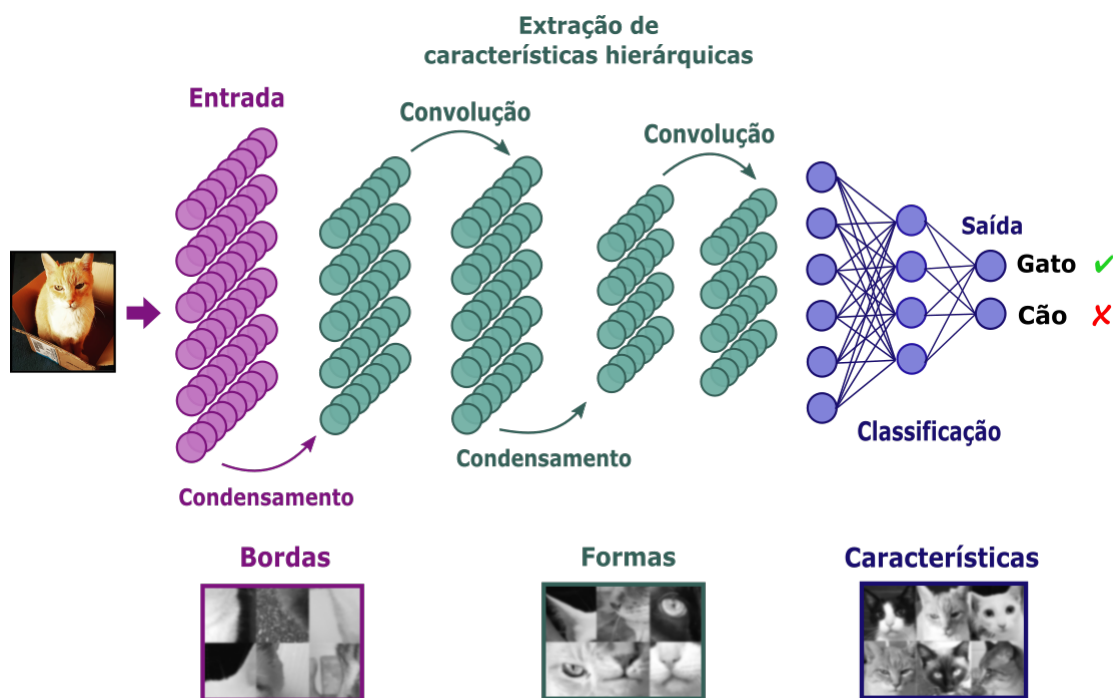


Figura 2: Esquema de uma rede convolucional simples.

Uma das arquiteturas mais comuns de redes convolucionais consiste em algumas camadas convolucionais e camadas de condensamento intercaladas. O papel destas camadas é extrair os descritores dos dados de entrada, enquanto que as últimas camadas servem como classificadores. Estas últimas costumam ser camadas simples de fileiras de neurônios. Por conta do viés

hierárquico, nas camadas mais iniciais é possível que se usem até mesmo pesos otimizados em bases de dados externas (*transfer learning*), que sejam significativamente maiores, já que estas primeiras camadas seriam responsáveis por identificar as características de mais baixo nível, isto é, de caráter mais universal [5].

3 Metodologia matematico-computacional

As últimas camadas de uma rede convolucional profunda também são conhecidas como “classificadores”. Elas recebem os descritores das imagens extraídos nas camadas convolucionais e determinam a classificação mais provável para cada imagem de teste. Existem porém diversos outros métodos que podem ser usados para fazer esta classificação, além das últimas camadas totalmente conectadas. Alguns deles são:

- **Fully Connected Layer (FCL):** trata-se do classificador usualmente utilizado na última camada da rede, é uma fileira simples de neurônios.
- **Support Vector Machine (SVM):** o objetivo do algoritmo é encontrar um hiperplano em um espaço N -dimensional que melhor separe os pontos que representam os dados [9]. Na Figura 3, vê-se um exemplo de um classificador SVM linear, em que é possível separar completamente as amostras. Pode-se observar o esquema dos hiperplanos e suas respectivas equações.

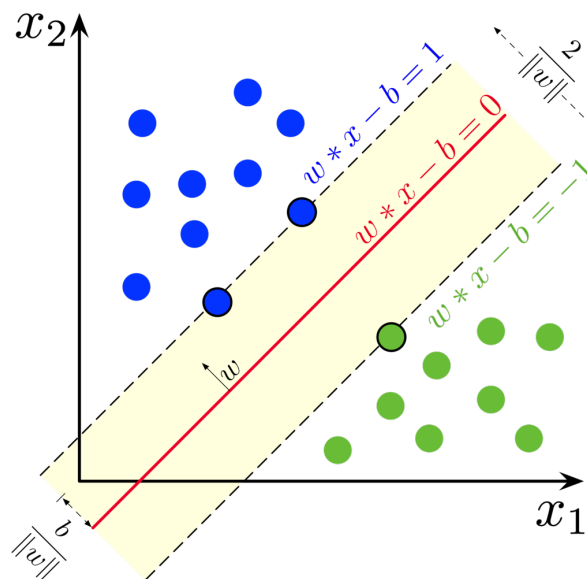


Figura 3: Esquema de um classificador SVM linear separável.

- **K-Nearest Neighbors (KNN):** consiste em verificar a classe dos K descritores mais próximos ao descritor da imagem a ser classificada (esta distância pode ser a Euclidiana ou qualquer outra) e classificar 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):** baseado no conceito de procurar uma combinação linear de variáveis que melhor separe duas classes. Trata-se de um método estatístico que

utiliza parâmetros como expectativa e covariância [10]. O método é dado pelas seguintes equações:

$$\vec{w} \cdot \vec{x} > c$$

$$\vec{w} = \Sigma^{-1}(\vec{\mu}_1 - \vec{\mu}_0)$$

$$c = \frac{1}{2}(T - \vec{\mu}_0^T \Sigma^{-1} \vec{\mu}_0 + \vec{\mu}_1^T \Sigma^{-1} \vec{\mu}_1),$$

em que \vec{x} é o vetor das amostras, $\vec{\mu}$ é o vetor das expectativas, Σ é a matriz de covariância e T é um limitante qualquer.

A metodologia utilizada neste projeto consiste em aplicar os descritores extraídos pelas camadas convolucionais aos diferentes classificadores citados acima. Cada classificador determinou a qual classe cada imagem de teste pertence. A partir disso, foi feito um sistema de votação, em que os classificadores em todos os conjuntos possíveis são agrupados, dois a dois, três a três, e assim por diante. Então faz-se a votação para cada um desses grupos de classificadores, escolhendo-se aquele que tiver a maior taxa de acerto. Em seguida ainda são excluídos do grupo o classificador que teve o pior desempenho, fazendo a votação final com os classificadores restantes.

O processo de votação é conhecido como comitê de máquinas (*ensemble* em inglês) e visa aumentar a acurácia da rede neural, por meio da comparação dos resultados de diferentes classificadores.

4 Aplicações em imagens biológicas

4.1 Identificação de espécies de plantas a partir de imagens da superfície foliar

O processo de identificação da espécie de um organismo vegetal é de grande importância tanto para especialistas de áreas como botânica e agricultura, quanto para o público leigo. Em termos práticos, uma identificação precisa permite um melhor controle de parâmetros como quantidade e tipo de nutrientes e corretores de solo a serem aplicados, além de outras condições passíveis de serem reguladas, assim aumentando a produtividade de uma safra ou a acurácia de um processo de melhoramento genético, por exemplo.

Assim como na abordagem convencional, a identificação automatizada de espécies também se baseia na análise de características-chave presentes na planta e que permitem a distinção entre as diferentes espécies. Classicamente, especialistas costumam focar esta análise em flores e frutos, por estes serem órgãos com características fáceis de serem observadas. Ocorre entretanto que estes órgãos não estão facilmente disponíveis em todas as plantas nem em todas as épocas do ano. Como as folhas estão disponíveis em qualquer época, especialistas passaram a buscar informações nessa parte da planta.

O problema que surge é que as folhas são um dos órgãos vegetais mais heterogêneos tanto anatomicamente quanto morfologicamente, 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. Estes

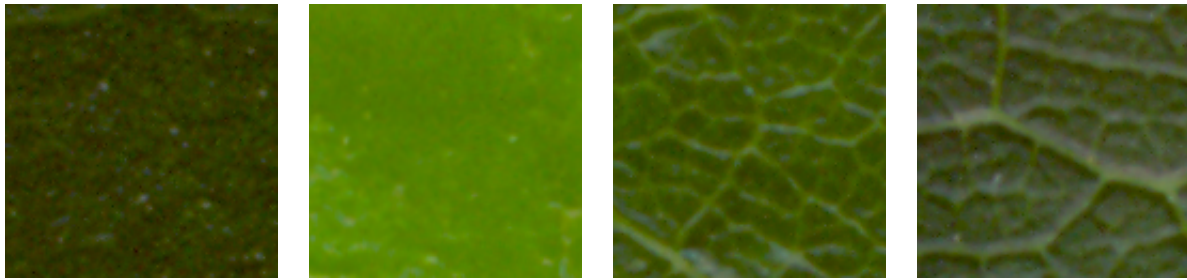


Figura 4: Exemplos de superfícies foliares analisadas.

métodos são capazes de analisar tanto estruturas externas quanto internas da folha e descrever com precisão atributos que não podem ser inferidos pelo olho humano, como padrões geométricos, texturais e de cores por exemplo.

O presente projeto aplicou a abordagem computacional desenvolvida combinando redes convolucionais com *ensembles* de classificadores na identificação de espécies de plantas brasileiras, mais especificamente da base 1200Tex proposta e estudada em [2]. Algumas amostras desta base são ilustradas na Figura 4.

4.2 Categorização de cistos bucais

Cistos odontogênicos são uma lesão da mandíbula, recentemente reclassificados pela Organização Mundial da Saúde como carcinogênicos [3]. Podem ser divididos em três grupos: radiculares, queratocistos esporádicos e queratocistos sindrômicos. A Figura 5 ilustra imagens de algumas amostras destes cistos. Seu estudo é de grande importância sobretudo devido a suas altas taxas de crescimento e recidiva, o que faz com que a identificação do tipo de cisto seja fundamental.

Embora atributos histológicos que diferenciem tais cistos sejam bem conhecidos por histopatologistas e permitam alguma análise por parte destes pesquisadores, o processo neste caso é manual, sujeito à subjetividade do diagnóstico feito por humanos, além de não permitir que a análise de um grande conjunto de dados seja possível em um intervalo de tempo razoável.

Para preencher esta lacuna, métodos computacionais de análise de imagens microscópicas vêm sendo aplicados na literatura nos últimos anos, alcançando resultados promissores. Em particular, as redes neurais convolucionais profundas se apresentam como uma abordagem bastante promissora neste tipo de problema [6].

Neste sentido, os descritores extraídos pela metodologia desenvolvida foram empregados na identificação e discriminação de cistos bucais com base nas imagens microscópicas coletadas a partir da biópsia de tecidos afetados.

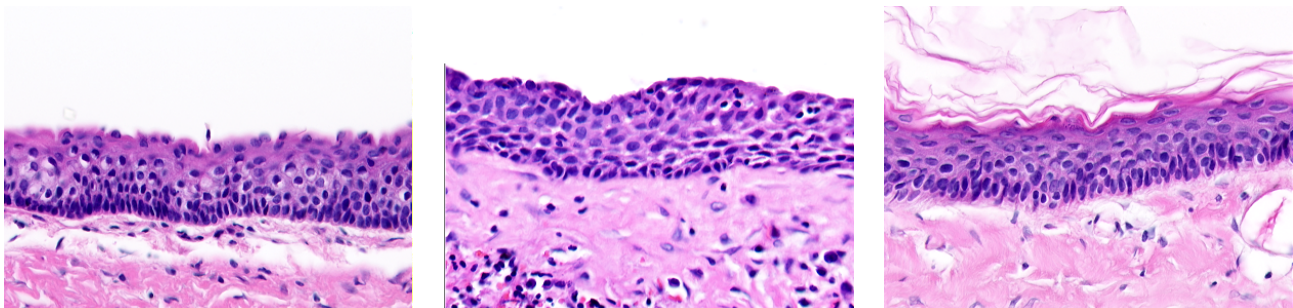


Figura 5: Imagens ilustrativas de algumas amostras dos cistos bucais estudados neste projeto.

4.3 Categorização de texturas sob condições distintas de iluminação, posição e escala

Um dos problemas em visão computacional que pode ser tratado por redes neurais é a classificação de imagens independentemente das condições sob as quais as fotos foram tiradas. Sabe-se que fatores externos podem alterar drasticamente as características de uma fotografia: a iluminação do ambiente, o ângulo com que a luz incide no material, a distância e a posição do material em relação à câmera, entre inúmeros outros aspectos. No entanto, espera-se que a rede seja capaz de classificar a imagem corretamente, mesmo submetida a essas alterações.

Para analisar esse aspecto da rede neural utilizada neste projeto, utiliza-se a base de dados conhecida como “KTH-TIPS2-2b” [7] (*Textures under varying Illumination, Pose and Scale*), que consiste em um conjunto de fotos de texturas provenientes de 11 materiais/classes distintos, sendo eles: folha de alumínio, pão integral, veludo, cortiça, algodão, biscoito, folha de alface, linho, pão branco, madeira e lã. Cada classe, por sua vez, é dividida em quatro subclasses nas quais as fotos foram submetidas a variações na iluminação, posição ou escala.

A metodologia utilizada consiste em escolher uma dessas amostras de cada classe e usar apenas esse conjunto de imagens para treinar a rede, enquanto as outras três subclasses são usadas para testar a rede. Com isso, é possível analisar qual é o resultado da rede ao se deparar com conjuntos de teste cujas imagens estão submetidas a condições de iluminação, posição ou escala à qual ela não tinha sido exposta anteriormente. O processo é repetido até que todas as subclasses tenham sido utilizadas uma vez como conjunto de treino da rede.

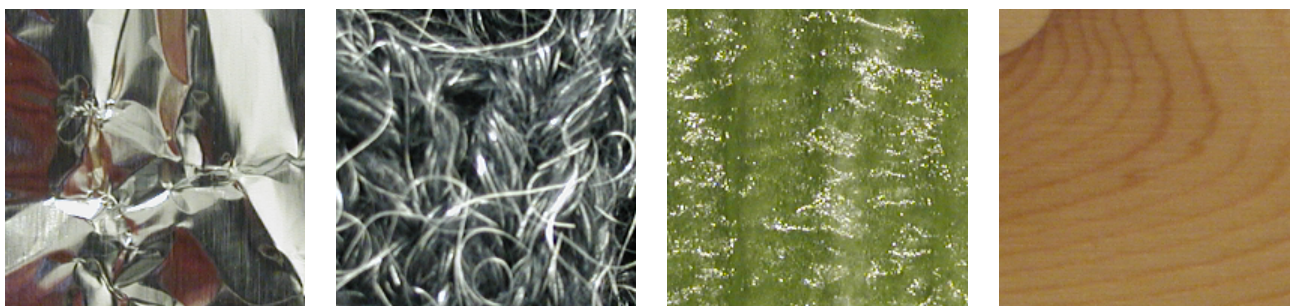


Figura 6: Exemplos de texturas analisadas.

5 Resultados e Análises

Todas as redes foram implementadas em *Python*, usando como base os tutoriais de “*Transfer Learning*” da biblioteca do *Pytorch*.

5.1 Identificação de espécies de plantas

Os conjuntos de teste e treino foram separados aleatoriamente, numa proporção de metade das fotos para o teste e metade para o treino, para cada uma das 20 classes. Abaixo constam os resultados aproximados da acurácia máxima usando as abordagens de ajuste fino e extração fixa de características, assim como as redes de pré-treinamento, Resnet e AlexNet. Para cada método, o experimento foi repetido dez vezes, então foi tirada a média e calculado o desvio padrão.

Os resultados estão apresentados na tabela abaixo, para a qual utiliza-se a seguinte legenda:

- **FCL:** Fully Connected Layer
- **LDA:** Linear Discriminant Analysis
- **SVM:** Support Vector Machine
- **RF:** Random Forest
- **KNN:** K-Nearest Neighbors
- **ENS:** Ensemble

Classificador	Resnet (learning rate = 0,0001)				Alexnet (learning rate = 0,0005)			
	Ajuste fino		Extração fixa		Ajuste fino		Extração fixa	
FCL	87,4%	1,4%	74,6%	1,3%	82,6%	0,7%	82,2%	1,2%
LDA	78,7%	1,9%	46,5%	2,1%	69,7%	1,7%	60,8%	1,8%
SVM	88,2%	1,0%	80,1%	0,9%	73,0%	1,8%	69,0%	1,9%
RF	76,1%	2,0%	55,6%	2,1%	56,1%	4,0%	51,6%	1,4%
KNN	83,9%	1,2%	75,3%	1,6%	66,6%	1,2%	59,5%	1,4%
ENS	87,3%	0,9%	77,5%	1,0%	72,7%	1,6%	67,7%	1,6%

Figura 7: Acurácia máxima média com seu respectivo desvio padrão ao lado.

Vale ressaltar que os classificadores *Fully Connected Layer* (FCL) e *Support Vector Machine* (SVM) foram os que obtiveram os melhores resultados, enquanto que a estratégia de *ensemble* apenas conseguiu manter esses resultados. O desvio padrão se manteve dentro dos parâmetros esperados.

Foram utilizadas 25 épocas para cada experimento, o valor da função de custo foi armazenado a cada estágio. Tirando a média entre as dez repetições estatísticas, foram obtidos os gráficos do valor médio da função de custo por época. Na Figura 8 são apresentados dois exemplos de gráficos gerados para duas configurações específicas da rede.

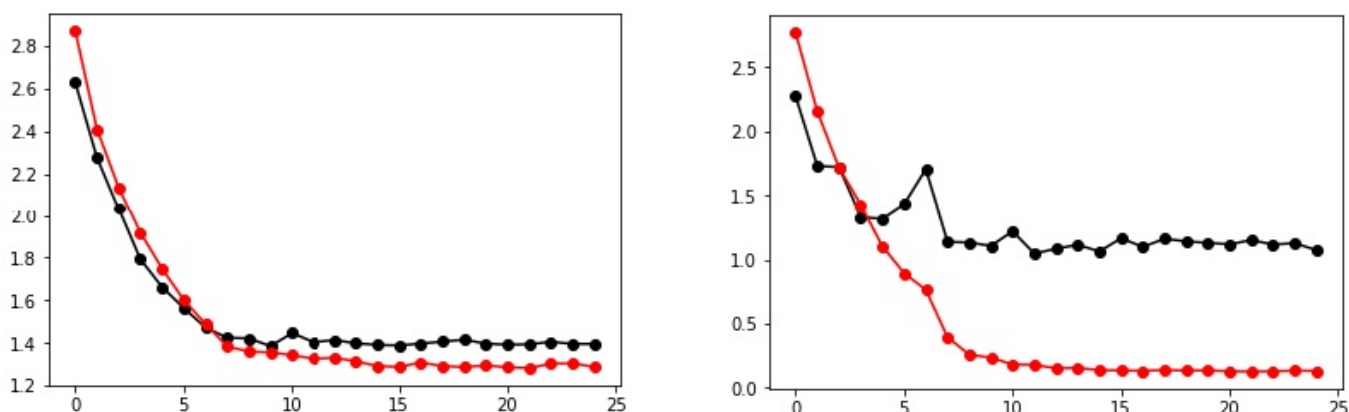


Figura 8: Gráfico do valor da função de custo por época (treino: vermelho, teste: preto).

O comportamento dos gráficos foi condizente com o esperado. Conforme a rede otimiza os pesos, a função de custo vai diminuindo até chegar a uma estabilidade, o que significa que a rede atingiu aproximadamente sua acurácia máxima.

Outra forma de visualizar os resultados é através da matriz de confusão, que mostra graficamente o número de imagens atribuídas a cada classe. Assim como no caso anterior, o resultado condiz com o esperado: uma diagonal ressaltada, que mostra que a rede classificou corretamente a maior parte das amostras.

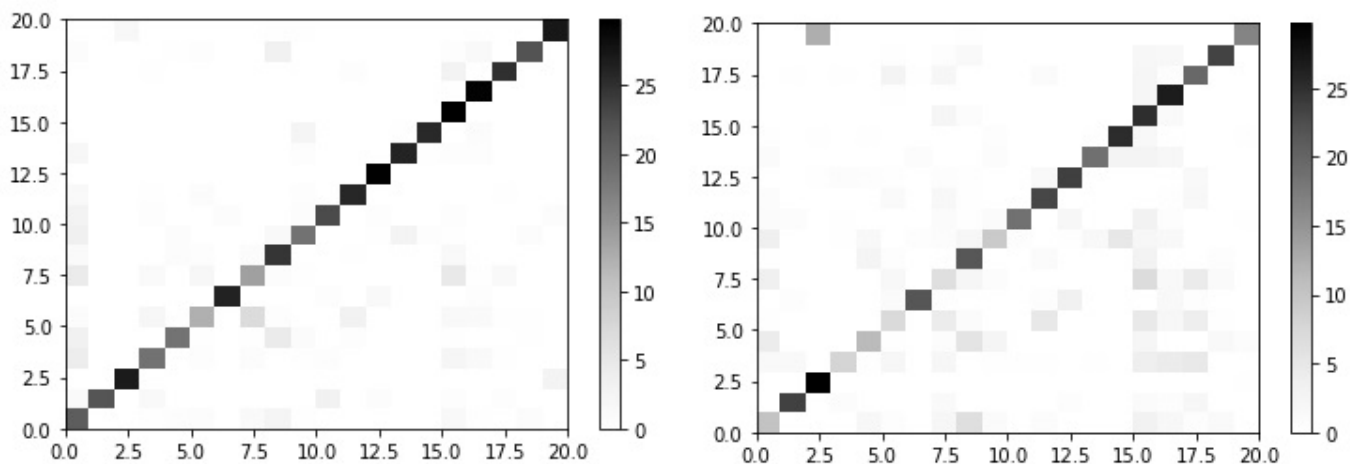


Figura 9: Representação em escala de cores das matrizes de confusão.

É notável que o método do ajuste fino apresentou resultados mais satisfatórios, o que era esperado, considerando-se que o tipo de imagem analisada é muito específico e, portanto, a base de dados externa ImageNet não teria tanto a acrescentar no desempenho da nossa rede. Por essa razão, quando todos os pesos são congelados, exceto os da última camada, obtém-se um resultado inferior comparado a quando apenas se inicializa a rede usando-se a base de dados externa e deixa-se que os pesos sejam otimizados a partir dos dados internos.

As matrizes de confusão evidenciam ainda mais a diferença entre a performance das duas metodologias. Na primeira, pode-se ver uma clara predominância nos valores da diagonal principal, ou seja, no número de imagens classificadas corretamente, sendo que não há valores significativos fora da diagonal. Na segunda, ainda é possível notar uma predominância na diagonal, mas consideravelmente menos acentuada, além de ser possível encontrar valores significativos fora dela.

5.2 Categorização de cistos bucais

Os conjuntos de teste e treino foram separados aleatoriamente, numa proporção de metade das fotos para o teste e metade para o treino, para cada uma das classes. Abaixo constam os resultados aproximados da acurácia máxima usando as abordagens de ajuste fino e extração fixa, e as redes de pré-treinamento, Resnet e AlexNet. Para cada método, o experimento foi repetido dez vezes, então foi tirada a média e calculado o desvio padrão.

Para esta base, além das variações já utilizadas na base de plantas, será repetido o experimento excluindo a classe "Radiculares".

Abaixo exibe-se um exemplo de matriz de confusão, que mostra o desempenho da rede para uma configuração específica.

$$\begin{bmatrix} 27,1 & 0,7 & 5,2 \\ 1,7 & 19,9 & 1,4 \\ 9,4 & 1,5 & 9,1 \end{bmatrix}$$

	Classificador	Sem radiculares (learning rate = 0,0001)				Com radiculares (learning rate = 0,0005)			
		Ajuste fino		Extração fixa		Ajuste fino		Extração fixa	
ResNet	FCL	77,4%	4,4%	68,5%	4,9%	75,0%	2,9%	75,8%	4,5%
	LDA	57,0%	9,4%	56,2%	5,8%	45,0%	4,1%	33,3%	8,8%
	SVM	81,9%	4,3%	81,3%	5,8%	79,5%	2,8%	79,2%	3,8%
	RF	70,6%	3,4%	69,8%	4,5%	70,0%	3,0%	66,4%	3,0%
	KNN	78,5%	6,5%	78,5%	4,7%	75,3%	4,4%	76,6%	3,5%
	RECALL	80,0%	0,3%	82,0%	0,2%	44,0%	2,0%	37,0%	1,7%
	ENS	82,1%	3,2%	81,5%	3,9%	79,5%	2,7%	80,0%	2,6%
AlexNet	FCL	78,7%	3,2%	76,0%	4,1%	82,0%	2,3%	78,6%	2,9%
	LDA	55,1%	6,9%	53,6%	8,4%	42,8%	6,7%	34,3%	8,0%
	SVM	64,9%	4,4%	60,8%	4,4%	65,9%	4,3%	62,1%	3,5%
	RF	68,7%	4,6%	67,4%	2,5%	64,1%	4,3%	64,1%	4,4%
	KNN	65,1%	6,7%	63,6%	7,5%	61,8%	6,2%	61,2%	5,1%
	RECALL	59,0%	1,2%	64,0%	0,8%	38,0%	1,8%	41,0%	2,0%
	ENS	72,8%	2,3%	69,6%	2,0%	68,2%	5,0%	67,6%	2,8%

Figura 10: Acurácia máxima média com seu respectivo desvio padrão ao lado.

É possível observar que a ResNet teve um desempenho superior, ficando em torno de 80% de acurácia, enquanto que a AlexNet ficou em torno de 70%. Novamente é possível se observar um destaque para os classificadores *Fully Connected Layer* (FCL) e *Support Vector Machine* (SVM). Para esta base, a estratégia de *ensembles* conseguiu, em geral, atingir um resultado levemente superior ao do melhor classificador, o que mostra que houve um acréscimo nos acertos obtidos, apesar de não ter sido um aumento significativo.

Uma peculiaridade desta base, é a de que o número de amostras para cada classe não é igual. Esse desequilíbrio leva ao questionamento do quão precisa é a medida de número de acertos por si só, já que a rede pode estar simplesmente classificando todas as amostras como pertencentes à classe mais numerosa, e ainda assim estaria dando um resultado “bom” quanto ao número de acertos.

Para resolver esse problema, pode-se usar uma estimativa de *precision-recall*, que consiste basicamente em uma relação entre os valores de falsos-positivos e falsos-negativos para diferentes limitantes (*thresholds*), formando um gráfico. A área abaixo desse gráfico serve como uma medida muito mais confiável do desempenho da rede. Esse valor consta na tabela da Figura 10 como “RECALL”. Abaixo, na Figura 11, alguns exemplos desses gráficos são exibidos para duas configurações específicas da rede.

5.3 Categorização de texturas

Para cada classe da base “KTH-TIPS2-b”, existem quatro subclasses, denominadas “*sample a*”, “*sample b*”, “*sample c*” e “*sample d*”. Abaixo constam os resultados aproximados da acurácia máxima obtida usando cada uma dessas amostras e um arranjo aleatório como conjunto de treino da rede.

Pode-se ver que sempre que uma amostra foi usada como conjunto de teste, a taxa de acerto ficou em torno de 80%, enquanto que quando é usado um arranjo aleatório de imagens de todas as amostras, o desempenho se mostrou consideravelmente superior. Isso já era previsto, pois no segundo caso a rede já tinha tido contato com imagens similares (pertencentes à mesma

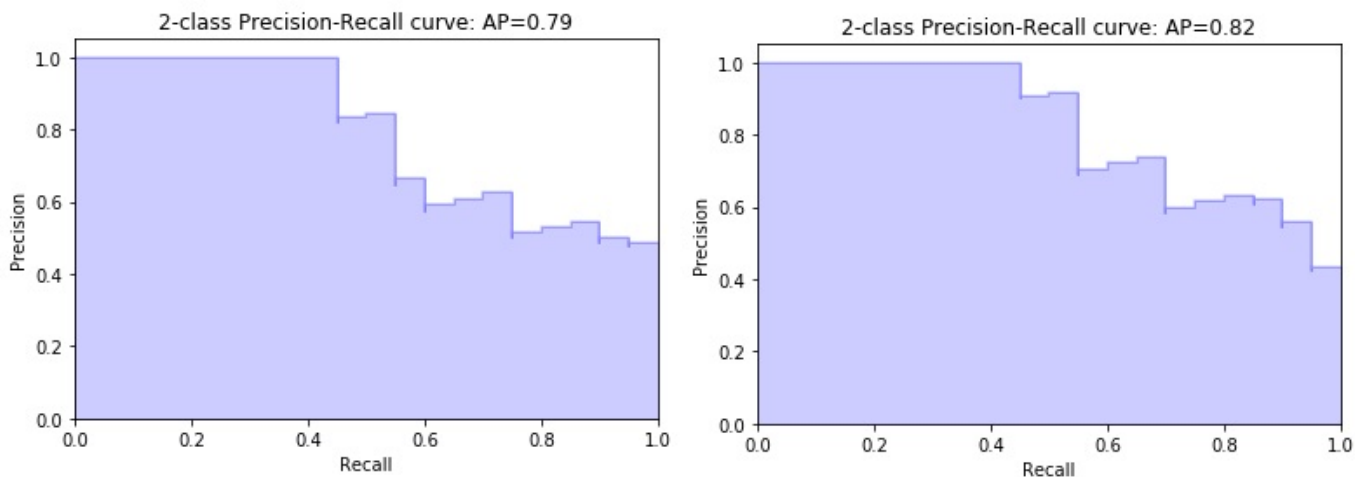


Figura 11: Gráficos de *precision-recall*.

	Classificador	Resnet (learning rate = 0,0001)				Alexnet (learning rate = 0,0005)			
		Ajuste fino		Extração fixa		Ajuste fino		Extração fixa	
Geral	FCL	98,9%	0,4%	91,3%	0,9%	99,7%	0,1%	98,5%	0,3%
	LDA	90,7%	2,5%	80,3%	2,2%	91,5%	3,0%	80,7%	3,4%
	SVM	99,8%	0,1%	99,3%	0,3%	97,3%	0,5%	92,8%	0,3%
	RF	94,7%	2,2%	75,7%	0,2%	81,6%	1,1%	67,4%	0,3%
	KNN	98,8%	0,4%	90,2%	0,9%	89,0%	0,6%	77,9%	0,8%
	ENS	99,4%	0,3%	95,6%	0,7%	93,6%	0,2%	85,8%	0,1%

Figura 12: Acurácia máxima média com seu respectivo desvio padrão ao lado.

amostra) durante o treino.

Esses resultados usando as amostras como conjunto de teste podem ser considerados satisfatórios, já que eles representam o desempenho da rede diante de imagens submetidas a condições adversas não vistas durante o treino. Isso significa que o método é 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, por exemplo.

Novamente é possível se observar que a ResNet obteve um resultado melhor do que a AlexNet, assim como os classificadores *Fully Connected Layer (FCL)* e *Support Vector Machine (SVM)*. A estratégia de *ensembles* conseguiu obter, novamente, resultados levemente superiores ao do melhor classificador.

	Classificador	ResNet (learning rate = 0,002)		AlexNet (learning rate = 0,0005)	
		Ajuste fino	Extração fixa	Ajuste fino	Extração fixa
Sample A	FCL	72,4%	68,4%	73,4%	75,4%
	LDA	69,0%	71,3%	65,3%	62,1%
	SVM	72,0%	73,9%	64,2%	64,4%
	RF	64,2%	58,9%	54,2%	50,0%
	KNN	71,5%	73,9%	66,8%	62,1%
	ENS	71,5%	74,9%	66,3%	64,3%
Sample B	FCL	75,6%	71,6%	74,8%	72,5%
	LDA	70,6%	69,0%	59,9%	60,9%
	SVM	75,6%	71,8%	60,1%	59,2%
	RF	68,7%	62,0%	50,9%	47,3%
	KNN	72,7%	72,8%	60,9%	57,9%
	ENS	75,2%	72,8%	61,3%	60,6%
Sample C	FCL	68,4%	63,1%	73,5%	73,3%
	LDA	69,4%	64,5%	58,3%	56,7%
	SVM	70,1%	67,5%	61,2%	58,6%
	RF	63,3%	58,6%	56,5%	52,0%
	KNN	61,3%	69,1%	60,2%	57,0%
	ENS	70,6%	69,7%	61,8%	59,5%
Sample D	FCL	76,7%	72,3%	69,1%	71,7%
	LDA	69,0%	67,8%	57,7%	55,0%
	SVM	73,8%	70,5%	56,6%	60,7%
	RF	54,5%	58,7%	44,9%	54,2%
	KNN	72,7%	73,0%	59,3%	62,9%
	ENS	73,5%	70,9%	59,5%	61,2%

Figura 13: Acurácia máxima média com seu respectivo desvio padrão ao lado.

Como o comportamento em geral dos gráficos da função de custo mostrou uma estabilização depois da décima época, reduziu-se o número de épocas de 25 para 15. Na Figura 13, já pode ser visto o uso dessa nova abordagem.

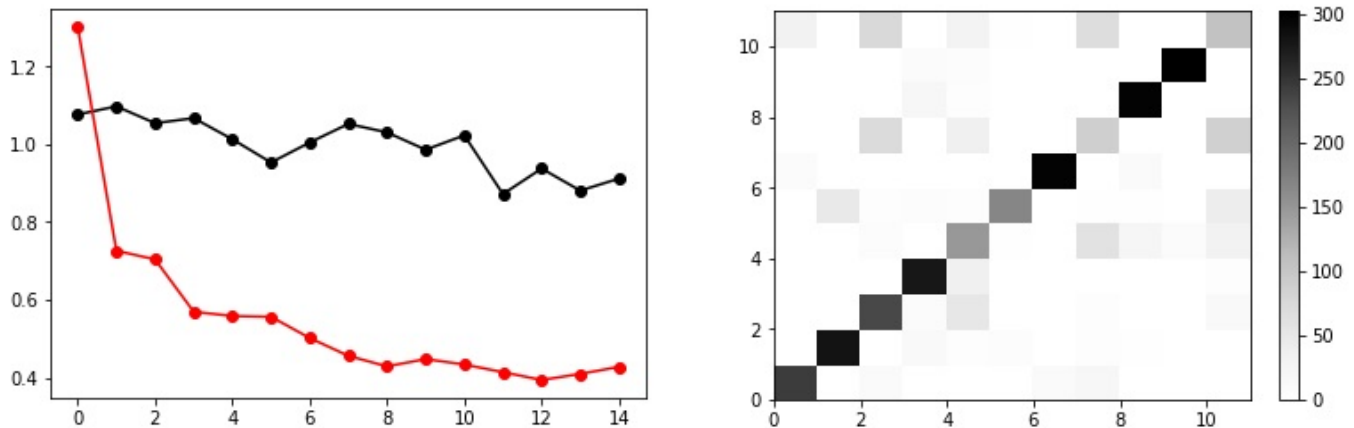


Figura 14: Representação em escala de cores das matrizes de confusão.

6 Conclusão

O projeto abordou o desenvolvimento científico e tecnológico em visão computacional e aprendizado de máquinas, bem como a aplicação na área médica e na botânica. Analisou-se a eficiência das redes neurais em categorização de imagens sob condições variadas, mostrando que a rede apresenta um desempenho satisfatório mesmo diante de condições adversas não observadas no processo de treinamento.

A premissa de extrair características da rede e utilizar diferentes classificadores se mostrou eficiente e funcional, evidenciando os classificadores *Fully Connected Layer* (FCL) e *Support Vector Machine* (SVM) como aqueles que obtiveram os melhores resultados para classificação de imagens. Já a estratégia de *ensembles* conseguiu atingir um desempenho tão bom quanto ou levemente superior ao dos melhores classificadores.

Referências

- [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] Ian Goodfellow, Yoshua Bengio, and Aaron Courville. *Deep Learning*. MIT Press, 2016. <http://www.deeplearningbook.org>.
- [5] Yann LeCun, Yoshua Bengio, and Geoffrey Hinton. Deep learning. *nature*, 521(7553):436, 2015.
- [6] 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.

- [7] 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.*
- [8] Juergen Schmidhuber, Xi Zhen-yuan, Wang Chang-li, Chen He, and Wu Jing. Deep learning in neural networks: An overview. In *2015 INTERNATIONAL CONFERENCE ON COMPUTERS, COMMUNICATIONS, AND SYSTEMS (ICCCS)*, pages 85–117, 2015.
- [9] Bernhard Scholkopf and Alexander J Smola. *Learning with kernels: support vector machines, regularization, optimization, and beyond.* MIT press, 2001.
- [10] Jieping Ye, Ravi Janardan, and Qi Li. Two-dimensional linear discriminant analysis. In *Advances in neural information processing systems*, pages 1569–1576, 2005.