UNIVERSIDADE ESTADUAL DE CAMPINAS

Instituto de Matemática, Estatística e Computação Científica

ACACIO NECKEL

Máquinas de aprendizado extremo aplicadas ao reconhecimento de imagens de texturas: uma aplicação em imagens médicas

Campinas
2020

Acacio Neckel

Máquinas de aprendizado extremo aplicadas ao reconhecimento de imagens de texturas: uma aplicação em imagens médicas

Dissertação apresentada ao Instituto de Matemática, Estatística e Computação Científica da Universidade Estadual de Campinas como parte dos requisitos exigidos para a obtenção do título de Mestre em Matemática Aplicada.

Orientador: João Batista Florindo

Este exemplar corresponde à versão final da Dissertação defendida pelo aluno Acacio Neckel e orientada pelo Prof. Dr. João Batista Florindo.

Campinas

2020

Ficha catalográfica
Universidade Estadual de Campinas
Biblioteca do Instituto de Matemática, Estatística e Computação Científica
Ana Regina Machado - CRB 8/5467

N282m

Neckel, Acacio, 1991-
Máquinas de aprendizado extremo aplicadas ao reconhecimento de imagens de texturas : uma aplicação em imagens médicas / Acacio Neckel. – Campinas, SP : [s.n.], 2020.

Orientador: João Batista Florindo.
Dissertação (mestrado) – Universidade Estadual de Campinas, Instituto de Matemática, Estatística e Computação Científica.

1. Máquinas de aprendizagem extremo. 2. Reconhecimento de textura. 3. Descritores locais. 4. Imagens médicas. I. Florindo, João Batista, 1984-. II. Universidade Estadual de Campinas. Instituto de Matemática, Estatística e Computação Científica. III. Título.

Informações para Biblioteca Digital

Título em outro idioma: Extreme learning machines applied to texture image recognition : an aplication in medical images
Palavras-chave em inglês:
Extreme learning machines
Texture recognition
Local descriptors
Medical images
Área de concentração: Matemática Aplicada
Titulação: Mestre em Matemática Aplicada
Banca examinadora:
João Batista Florindo [Orientador]
André Ricardo Backes
Jarbas Joaci de Mesquita Sá Junior
Data de defesa: 01-09-2020
Programa de Pós-Graduação: Matemática Aplicada

Identificação e informações acadêmicas do(a) aluno(a)
- ORCID do autor: https://orcid.org/0000-0002-4998-0498
- Currículo Lattes do autor: http://lattes.cnpq.br/5636876436703297

Dissertação de Mestrado defendida em 01 de setembro de 2020 e aprovada pela banca examinadora composta pelos Profs. Drs.

Prof(a). Dr(a). JOÃO BATISTA FLORINDO

Prof(a). Dr(a). ANDRÉ RICARDO BACKES

Prof(a). Dr(a). JARBAS JOACI DE MESQUITA SÁ JUNIOR

A Ata da Defesa, assinada pelos membros da Comissão Examinadora, consta no SIGA/Sistema de Fluxo de Dissertação/Tese e na Secretaria de Pós-Graduação do Instituto de Matemática, Estatística e Computação Científica.

Agradecimentos

Ao professor João Batista Florindo pela orientação, ensinamentos, paciência, dedicação e colaboração no desenvolvimento deste trabalho.

À minha família e amigos, pelo apoio durante os estudos.

Aos professores e colaboradores do Instituto de Matemática, Estatística e Computação Científica pelos ensinamentos na trajetória do curso.

À UNICAMP pelo auxílio estudantil e a disponibilização do curso.

O presente trabalho foi realizado com apoio da Coordenação de Aperfeiçoamento de Pessoal de Nível Superior – Brasil (CAPES) – Código de Financiamento 001.

Resumo

A classificação de imagens é uma das tarefas de destaque na visão computacional. Entre os tipos de imagens que possuem grande aplicação na prática estão as imagens de texturas, nas quais todos os pixels apresentam igual influência no reconhecimento da imagem. Algoritmos computacionais para reconhecimento de texturas têm sido amplamente investigados na literatura, tanto para aumentar a acurácia e confiabilidade do processo de classificação quanto para reduzir o tempo computacional. A proposta apresentada neste trabalho visou desenvolver descritores locais de imagens de texturas por meio de um processo baseado na teoria de aprendizado automático, mais especificamente, empregando máquinas de aprendizado extremo. A metodologia proposta explora o uso dessas máquinas tanto no domínio original da imagem quanto sobre uma transformação baseada em vizinhança. Propõe-se também um método na direção inversa, isto é, usando máquinas de aprendizado extremo para formular uma transformada da imagem seguindo-se da extração de descritores locais sobre essa transformada. Outra solução explorada é o uso de uma operação alternativa para o produto interno associado ao algoritmo de aprendizado extremo, baseando-se em operadores morfológicos da álgebra min-max. Os algoritmos desenvolvidos foram aplicados em imagens médicas, na identificação de subtipos de câncer de pulmão. Os resultados obtidos tanto na aplicação médica quanto em bases de imagens de *benchmark* mostraram-se promissores, confirmando um aumento na acurácia da classificação em várias situações.

Palavras-chave: Máquinas de Aprendizado Extremo, Reconhecimento de Texturas, Descritores Locais, Imagens Médicas.

Abstract

Image classification is one of the most important tasks in computer vision. An example of images that have great application in practice are texture images, in which all pixels have an equal influence for image recognition purposes. Computational algorithms for texture recognition have been extensively investigated in the literature, both to increase the accuracy and reliability of the classification process as well as to reduce computational time. The proposal presented in this work aimed to develop local textural image descriptors through a process based on automatic learning theory, more specifically, using extreme learning machines. The proposed methodology explores the use of these machines both in the original image domain and in a neighborhood-based transformation. A method is also proposed in the reverse direction, that is, using extreme learning machines to formulate an image transform followed by the extraction of local descriptors over that transform. Another solution explored here is the use of an alternative operation for the inner product associated with the extreme learning algorithm, based on morphological operators of the min-max algebra. The developed algorithms were applied to medical images, in the identification of lung cancer subtypes. The results obtained both in the medical application and in benchmark databases were promising, confirming an increase in the accuracy of the classification in several situations.

Keywords: Extreme Learning Machines, Texture Recognition, Local Descriptors, Medical Images.

Sumário

1 Introdução

A visão computacional (SZELISKI, 2010) é uma das áreas em destaque da ciência moderna, fomentando assim um grande número de estudos na literatura, tanto teóricos quanto de novas formas e áreas de aplicação, que vão desde o reconhecimento de padrões em imagens médicas (SENGUPTA et al., 2020) até carros autônomos (ZHOU et al., 2019), entre muitos outros.

Dentro da visão computacional, um tipo de problema que tem grande interesse prático é o reconhecimento de imagens de texturas. Estas imagens, também chamadas de texturas visuais, não possuem uma definição formal de consenso na literatura, mas podem ser consideradas intuitivamente como sendo aquelas em que todos os pixels contribuem de maneira igualitária para sua análise, ou ainda, em que a distribuição estatística de padrões destes pixels é o fator preponderante em sua representação. Embora o contexto original dessas imagens esteja frequentemente associado a amostras de materiais, os algoritmos desenvolvidos para essas tarefas encontram hoje um vasto campo de aplicação. Imagens médicas microscópicas, sobretudo de tecidos ou núcleos celulares, são um bom exemplo de cenário em que técnicas de análise de texturas costumam ser efetivas.

Como visa automatizar tarefas usualmente executadas por humanos através de seu sistema visual, a visão computacional acaba sendo uma área de estudo bastante complexa e multidisciplinar, envolvendo conceitos da computação, da matemática, da estatística, da física, etc. A grosso modo, pode-se dizer que um sistema de visão computacional comumente combina técnicas de duas grandes áreas: análise de imagens e inteligência computacional.

Assim como a própria visão computacional, a inteligência computacional (ou artificial) é também uma área de desenvolvimento tecnológico que tem tido grande destaque nos últimos anos. Os algoritmos que mais têm se destacado nesta área são aqueles baseados na chamada “teoria do aprendizado automático” (HASTIE et al., 2001), em que existe um processo de treinamento no qual o algoritmo “aprende” a executar tarefas complexas a partir do *feedback* dado por um agente externo (normalmente um especialista humano) para aquela mesma tarefa. Entre os métodos nesta linha, as máquinas de aprendizado extremo (sigla ELM de *extreme learning machine* em inglês) (HUANG et al., 2006) têm chamado atenção na literatura, principalmente devido à sua alta eficiência computacional e pelo processo exato e de passo único (não iterativo) que é usado no treinamento, possibilitado pelo uso de uma ferramenta matemática conhecida da teoria de matrizes: pseudo-inversa de Moore-Penrose (GOLUB; LOAN, 2013).

Embora o algoritmo original dos ELMs possa ser empregado em classificação de imagens, o seu dado de entrada é unidimensional, o que exige que a imagem seja transformada em um vetor e deste modo perca a relação existente entre pixels vizinhos, informação esta que é sabidamente crucial, especialmente para a representação de texturas. Deste modo, parte-se neste trabalho da metodologia desenvolvida em (SÁ JUNIOR; BACKES, 2016), na qual um ELM é usado para extrair descritores locais da imagem e estes descritores podem então, por sua vez, ser classificados por métodos tradicionais como máquinas de vetores de suporte (VAPNIK, 2000), análise discriminante linear (HASTIE et al., 2001), florestas aleatórias (BREIMAN, 2001), etc.

A metodologia em (SÁ JUNIOR; BACKES, 2016) foca diretamente no pixel da imagem e usa o algoritmo clássico de redes neurais *feed-forward* adotado nas máquinas de aprendizado extremo, de modo que ela poderia se beneficiar de uma análise em um nível mais abstrato, em que a aplicação do algoritmo se dá sobre representações locais da imagem de textura, ou ainda de novas variantes do algoritmo de rede neural adotado.

Neste sentido, este estudo aplica uma metodologia baseada em (SÁ JUNIOR; BACKES, 2016), porém propondo algumas variantes. Primeiramente, o algoritmo ELM local não é aplicado diretamente sobre o pixel, mas sim sobre padrões locais de textura. Para este propósito, são usados aqui os padrões locais binários (sigla LBP de *local binary patterns* em inglês) (OJALA et al., 2002), tanto em função de sua simplicidade de implementação e interpretação quanto por seu baixo custo computacional. Outra proposta apresentada neste trabalho é a chamada transformação ELM, a qual vai na direção oposta da primeira abordagem. Assim, a rede

ELM treinada a partir da imagem original com um número reduzido de neurônios é usada para reconstruir essa imagem, porém, adicionando um grau de compressão, já que a reconstrução não é exata. Com isso, temos uma imagem que se assemelha à original no sentido de que ambas são geradas pela mesma rede ELM, mas, ao mesmo tempo, remove detalhes e ruídos desnecessários que poderiam comprometer a eficiência da generalização do algoritmo para um conjunto de imagens de teste. Outro estudo realizado neste projeto envolve o uso de neurônios morfológicos, como descrito em (SUSSNER; CAMPIOTTI, 2020). Por fim, uma análise sobre o impacto de um processo de regularização sobre o ELM foi também analisado para efeitos de reconhecimento de texturas.

Além de investigar a eficácia do método proposto na classificação de imagens de texturas e comparar com o desempenho de outros algoritmos no estado-da-arte, também foi realizado um estudo matemático do funcionamento destes descritores baseados em ELM. O fato de os ELMs serem baseados em uma operação matemática conhecida, direta e relativamente simples, permite, por exemplo, que estudos assintóticos acerca da evolução do algoritmo sejam realizados.

1.1 Objetivos

O algoritmo de ELM utilizado neste trabalho baseia-se em uma rede neural com três camadas: entrada, oculta e de saída (HUANG et al., 2006). Os parâmetros da camada de entrada são organizados em um vetor coluna para cada amostra (que compõem uma matriz X quando a base toda é analisada). Na ligação entre a camada de entrada com a oculta, cada elemento da primeira camada é conectada com todos os neurônios da segunda, possuindo esta ligação um fator multiplicativo chamado de peso (matriz W). De modo similar, existem pesos entre a camada oculta e a de saída (matriz M) e, por fim, a resposta dos neurônios da camada de saída são organizadas na matriz Y.

Na aplicação do algoritmo ELM em extração de vetores de características em (SÁ JUNIOR; BACKES, 2016), os autores apresentam a seguinte sequência: dada a matriz representando uma imagem digital, uma submatriz S , chamada de "máscara" e usualmente bem menor do que a imagem original, percorre essa imagem coletando dois tipos de dados, o valor do pixel central sob S, que é colocado na matriz Y (saída), e dos pixels em volta (com determinado raio e seguindo uma sequência pré-determinada), que são atribuídos então a um vetor coluna na matriz X (amostra). Com isso, é possível aplicar o algoritmo ELM sobre a imagem bidimensional e o vetor de pesos aprendidos pela rede corresponde ao vetor de descritores daquela imagem.

As metodologias propostas neste estudo baseiam-se em modificações deste algoritmo para extrair descritores da imagem, bem como na combinação entre diferentes possibilidades.

A primeira abordagem apresentada baseia-se na utilização de uma transformada aplicada sobre a imagem original antes da aplicação efetiva do algoritmo ELM, visando assim detectar novos padrões na relação entre os pixels. Neste trabalho, foram usados padrões locais binários (LBP) para que na sequência houvesse a aplicação do ELM resultando no vetor de descritores. Deste modo, tanto os dados de entrada do modelo ELM (matriz X) quanto de saída (vetor Y) foram modificados para conter os códigos LBP associados a cada pixel, em vez do valor direto do pixel como na solução original em (SÁ JUNIOR; BACKES, 2016).

Uma segunda variante proposta neste projeto é o que chamamos de "transformação ELM". Nesta, o algoritmo ELM em (SÁ JUNIOR; BACKES, 2016) é aplicado sobre a imagem gerando um vetor de pesos (descritores). Na sequência, o mesmo algoritmo é executado novamente sobre a imagem, porém agora os pesos aprendidos na etapa anterior dão origem a um novo vetor Y, que por sua vez é usado com o intuito de reconstruir a imagem original. Este procedimento possibilita que cada pixel da imagem reconstruída tenha influência de todos os pixeis vizinhos na imagem original. Ao final, são extraídos descritores LBP da imagem reconstruída. Note-se que esta segunda abordagem pode ser considerada como seguindo o caminho inverso da primeira: antes os descritores ELM eram calculados a partir dos códigos LBP, agora os descritores LBP são calculados a partir da transformada ELM.

Ainda na linha de combinações entre ELM e LBP, uma terceira abordagem consiste em usar descritores LBP apenas na saída da rede ELM (vetor Y), enquanto que as os vetores que compõem a matriz de entrada X (vizinhos do pixel de referência) continuam sendo obtidos diretamente a partir da imagem original.

Por fim, propõe-se a utilização de uma variante morfológica para a rede neural associada ao ELM. Neste caso, em vez da operação de soma de produtos, tipicamente executada em um neurônio clássico, usa-se uma operação de máximo (ou mínimo) de somas (ou produtos), operação essa que está embasada na chamada “álgebra min-max” e que generaliza operadores da morfologia matemática. Este procedimento tem sido utilizado como classificador na literatura (SUSSNER; CAMPIOTTI, 2020) e no presente trabalho aplicamos essa solução para extrair descritores de texturas.

Na Figura 1 temos um diagrama geral das metodologias apresentadas para extração de descritores de uma imagem de textura.

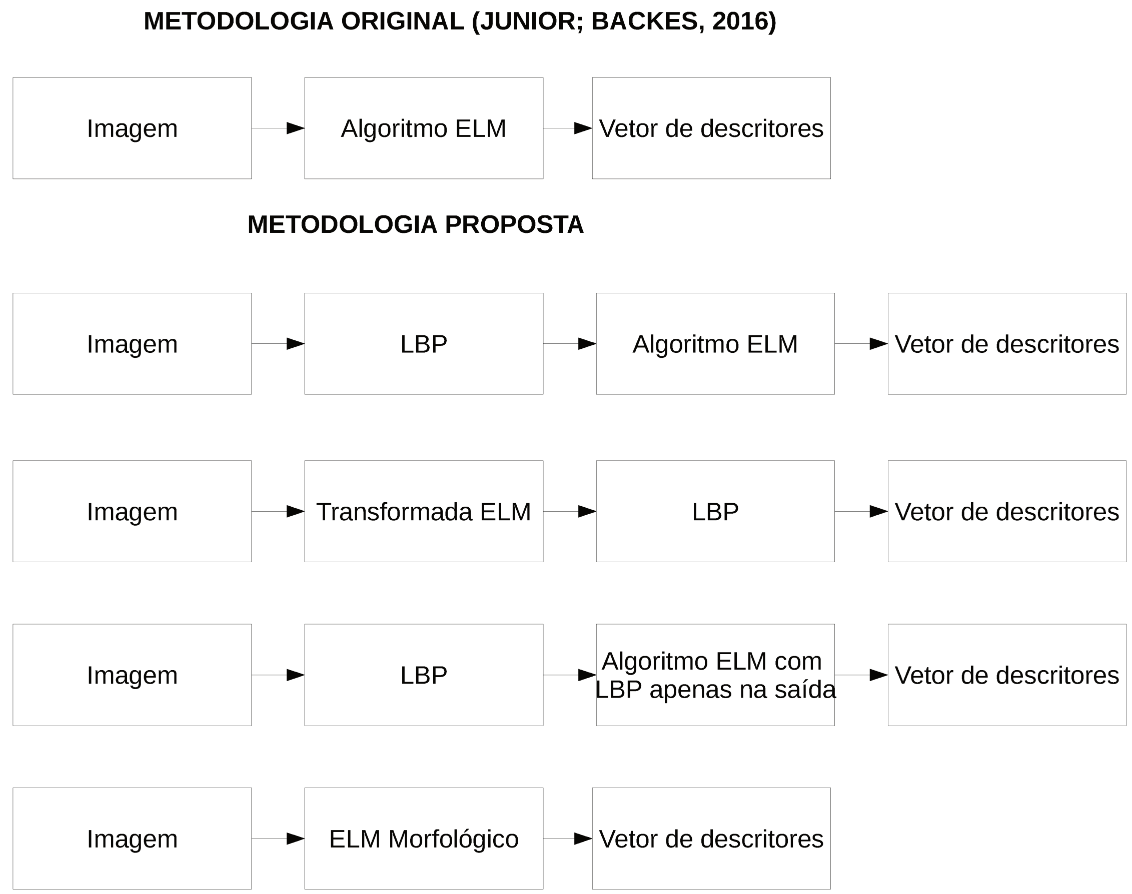


Figura 1 – Metodologia proposta para extração de descritores de imagens.

Para fins de comparação, as metodologias propostas foram testadas em três bases de imagens de *benchmark*, que possuem resultados de classificação amplamente difundidos na literatura: UMD (LAZEBNIK et al., 2005), UIUC (XU et al., 2009b) e

KTHTIPS-2b (HAYMAN et al., 2004).

Por fim, a metodologia proposta foi aplicada em um problema prático de grande interesse na área médica, qual seja o da identificação de tipos de câncer de pulmão por meio da imagem microscópica de cortes citológicos (núcleos celulares individuais).

2 Revisão Bibliográfica

Uma das abordagens que têm sido mais bem sucedidas na teoria de aprendizado de máquinas é a de redes neurais. Estas podem ser abstratamente caracterizadas por no mínimo três camadas. A primeira é uma camada de entrada, na qual os dados a serem analisados são inseridos no algoritmo. Em seguida existem uma ou várias camadas intermediárias (ocultas), nas quais são aplicados processamentos básicos baseados em uma soma ponderada do dado de entrada por parâmetros aprendidos pela rede, os chamados *pesos* e *biases*, e uma função de ativação não linear. Por fim, uma camada de saída é responsável por fornecer o resultado do processamento da rede.

Um tipo particular de rede neural é a chamada máquina de aprendizado extremo (HUANG et al., 2006). Esta costuma ter apenas três camadas: entrada, oculta e de saída. Esta abordagem se diferencia das demais, sobretudo, pelo fato de que os pesos que conectam a camada de entrada com a oculta são determinados aleatoriamente, enquanto os pesos conectando a camada oculta com a saída são obtidos por um cálculo exato e não iterativo (pseudo-inversa de Moore-Penrose).

As redes neurais, normalmente, são ferramentas de propósito geral e, deste modo, podem ser aplicadas a qualquer tipo de dado, incluindo imagens digitais. Sabe-se, entretanto, que o uso direto de uma rede convencional com imagens exige que a mesma seja transformada em um vetor, o que por sua vez acarreta perda de toda a estrutura local da imagem, que é de importância fundamental em tarefas de reconhecimento. Este comportamento fica ainda mais evidente nas imagens de texturas. Isto levou naturalmente ao desenvolvimento de adaptações da rede clássica para que fossem mais adequadas à análise de imagens, como é o caso das redes convolucionais (HAYKIN, 2001).

Um fenômeno parecido também ocorre com as máquinas de aprendizado extremo. Sua versão original tal como apresentada em (HUANG et al., 2006) não é eficiente para análise de texturas. Porém, a adaptação proposta em (SÁ JUNIOR; BACKES, 2016) mostrou-se bastante adequada para este propósito. Em (SÁ JUNIOR;

BACKES, 2016), o valor dos pixels da imagem são usados como saída esperada da rede ELM enquanto que os respectivos vizinhos são usados para compor a entrada. Os pesos aprendidos pelo processo de treinamento desta rede são então empregados como descritores da imagem original.

Estudos visando aumentar a eficiência do algoritmo ELM original bem como aplicá-lo em outros domínios de problemas têm sido apresentados na literatura. Por exemplo, uma adaptação do mesmo algoritmo de ELM apresentado em (SÁ JUNIOR; BACKES, 2016) é proposto em (SÁ JUNIOR et al., 2018) para a extração de vetores de características de imagens binárias de formas ou contornos, alcançando também excelentes resultados na classificação desses objetos.

Em (SÁ JUNIOR et al., 2019), os autores adaptaram o ELM para análise de texturas dinâmicas. Na abordagem proposta, a textura dinâmica é analisada em seus três planos de direção, cada um servindo como uma entrada independente para o algoritmo original. Ao final, os descritores obtidos são concatenados gerando um descritor geral. Os resultados encontrados comprovam a eficácia da metodologia proposta.

No trabalho (BACKES et al., 2019), os autores exploram e extração de vetores de características de imagens de textura com cor. Neste caso, os descritores dos canais R, G e B são concatenados para compor um descritor geral da imagem, que nos experimentos apresentados demonstram bom desempenho na tarefa de reconhecimento de texturas.

Modificações na maneira de calcular as entradas para a matriz X do modelo ELM também são apresentadas na literatura. Enquanto em (SÁ JUNIOR; BACKES, 2016) utiliza-se um sistema de submatrizes sobre a imagem, em (RIBAS et al., 2020) a entrada da rede é calculada com o apoio de uma rede complexa, diminuindo assim o tempo computacional para extrair os descritores de uma imagem e aumentando a porcentagem de acerto na classificação.

Em relação ao vetor Y, na metodologia proposta por (SÁ JUNIOR; BACKES, 2016), cada dado rotulado como saída é composto por apenas um único pixel, que pode ser o centro de uma janela 3×3, 5×5 ou 7×7. Já no trabalho de (JUNIOR; SÁ JUNIOR, 2019), para janelas maiores que 3×3, foi usado um conjunto de pixels como saída ou uma média entre eles. A primeira abordagem gera um descritor bidimensional, enquanto na segunda o descritor é unidimensional. As taxas de acerto na classificação são melhores do que a versão original do algoritmo.

Essas metodologias têm tido também grande sucesso em aplicações práticas, em problemas do mundo real. Por exemplo, em (SÁ JUNIOR et al., 2018b), o algoritmo de descritores ELM é aplicado à identificação de dois tipos de ligas de titânio. Já em imagens médicas, em (SÁ JUNIOR et al., 2018a) este tipo de solução é empregado na classificação do exame de Papanicolau em normal ou anormal. Em ambas as aplicações, a metodologia de ELM alcança os melhores resultados quando comparada com outros métodos da literatura.

Vale destacar também que o próprio algoritmo original de ELM (HUANG et al., 2006) tem sofrido transformações no decorrer do tempo na literatura. Por exemplo, no trabalho de (HUANG et al., 2017) o ELM é associado com campos receptivos locais e em (SUSSNER; CAMPIOTTI, 2020) é feita a associação de ELMs com redes neurais morfológicas, sendo que ambas as adaptações propiciaram aumento na acurácia do método.

Os mesmos padrões binários aqui explorados em combinação com o modelo ELM de (SÁ JUNIOR; BACKES, 2016) são também associados com o ELM clássico de (HUANG et al., 2006) em (TURKOGLU; HANBAY, 2019). Esse trabalho propõe o uso de uma modificação no LBP original, em que a comparação entre o pixel central e seus vizinhos se dá de diferentes maneiras. Estes procedimentos são utilizados para a extração de descritores da imagem de textura e na sequência o ELM tradicional de (HUANG et al., 2006) é empregado para classificação. Vale destacar aqui que apesar de se tratar também de uma combinação de ELM com LBP, a abordagem difere substancialmente da estudada neste mestrado, já que lá o ELM é um classificador e o LBP um descritor, enquanto aqui o ELM é um descritor e o LBP funciona como uma transformada da imagem.

Trazendo para um contexto mais voltado a aplicações, o reconhecimento de padrões em imagens médicas tem se mostrado uma tarefa em que o uso de ELMs pode ser interessante. O método é utilizado, por exemplo, para a classificação de câncer de pulmão como benigno ou maligno em (WAJID et al., 2016), a partir de radiografias do tórax, tendo obtido taxas de acerto acima de 90%. Em (WANG et al., 2018), foram utilizados cinco classificadores diferentes: máquina de vetores de suporte, rede neural probabilística, perceptron de múltiplas camadas, ELM e um algoritmo semi-supervisionado baseado no ELM (SS-ELM). As melhores taxas de acerto foram obtidas com SS-ELM e ELM.

Em (NAYAK et al., 2020), os autores trabalharam com ressonâncias magnéticas cerebrais. O conjunto de dados é composto por 200 imagens divididas em 5 classes: normal (controle), acidente vascular cerebral, tumor degenerativo, infeccioso e cerebral. Um ELM com duas camadas ocultas e a função de ativação *Leaky ReLU* é utilizado como classificador. Os autores comparam com outros sete métodos na literatura e obtêm as melhores taxas de acerto com ELM.

Similarmente, ainda na classificação de imagens médicas, mas agora usando ressonâncias magnéticas cerebrais, os autores em (LU et al., 2018) analisam amostras saudáveis e anormais com a utilização de um ELM modificado por um *kernel* específico e alcançam uma acurácia em torno de 97%.

Já em (MELEKOODAPPATTU; SUBBIAN, 2019), em um estudo de mamografias para identificar calcificação mamária usando um método de ELM híbrido, até 99% de acerto foi obtido.

Em suma, pode-se constatar que abordagens baseadas em ELM têm-se mostrado eficientes em problemas de visão computacional, tanto para extrair descritores de imagens quando para a classificação em si. Desde a proposta inicial do algoritmo em (HUANG et al., 2006), diferentes modificações têm contribuído cada vez mais tanto do ponto de vista teórico quanto em aplicações a diversos problemas do mundo real.

3 Máquinas de Aprendizado Extremo

Este capítulo apresenta o algoritmo básico das Máquinas de Aprendizado Extremo, discutido em (Pao; Takefuji, 1992; Schmidt et al., 1992; PAO et al., 1994; HUANG et al., 2006).

3.1 Descrição

O modelo de Máquinas de Aprendizado Extremo (ELM) apresentado em (HUANG et al., 2006) baseia-se em uma rede neural com uma camada de entrada, uma oculta e uma de saída. Uma visualização gráfica pode ser observada na Figura 2.

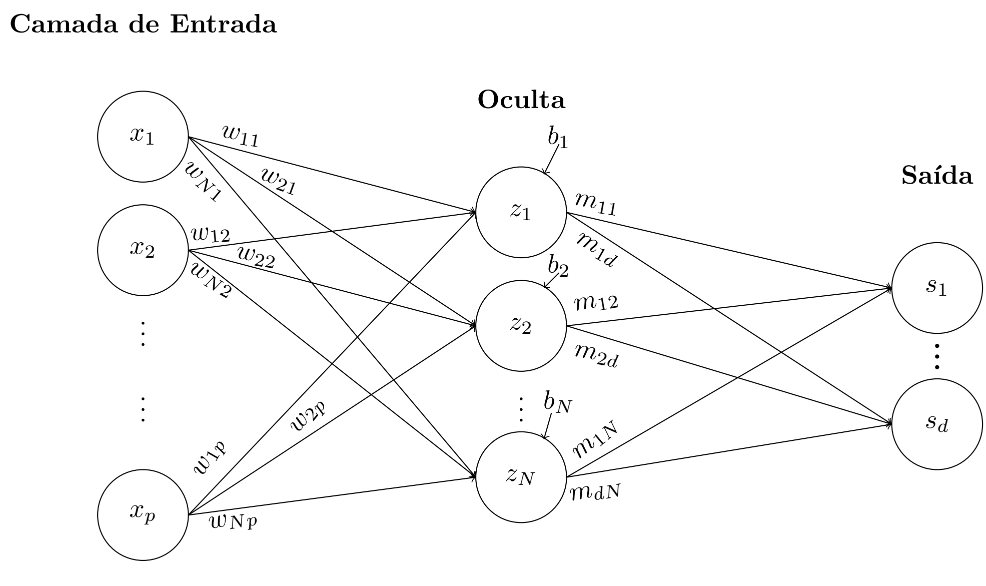


Figura 2 – Representação de um modelo para Máquina de Aprendizado Extremo.

Os pesos que conectam a camada de entrada com a camada oculta são gerados aleatoriamente, enquanto os pesos que conectam a camada oculta com a de saída são calculados pelo conceito matemático de mínimos quadrados (GOLUB; LOAN, 2013). Utilizando uma representação matricial e seguindo o caminho descrito por (SÁ JUNIOR; BACKES, 2016), os seguintes passos são executados.

Seja $X = [\mathbf{x}_1, \mathbf{x}_2, \cdots, \mathbf{x}_T]$ a matriz de entradas, com $\mathbf{x}_i$ representando cada vetor da amostra e pertencendo ao $\mathbb{R}^p$ e T o número total de amostras. A última linha da matriz X, é composta com todos os elementos iguais a 1, de modo que quando X é multiplicado pela matriz W, o valor do *bias* do neurônio correspondente já está sendo somado.

Seja ainda $Y = [\mathbf{y}_1, \mathbf{y}_2, \cdots, \mathbf{y}_T]$, com $\mathbf{y}_i$ representando a saída correspondente de cada vetor da amostra e pertencendo ao $\mathbb{R}^d$.

A matriz W, com todos os pesos entre a camada de entrada e oculta, pode ser representada por:

$$W = \begin{bmatrix} w_{11} & w_{12} & \cdots & w_{1p} & b_1 \\ w_{21} & w_{22} & \cdots & w_{2p} & b_2 \\ \vdots & \vdots & \vdots & \vdots & \vdots \\ w_{N1} & w_{N2} & \cdots & w_{Np} & b_N \end{bmatrix}, \tag{3.1}$$

em que N é o número de neurônios da camada oculta. Cada linha i da matriz W representa os respectivos pesos e o *bias* associados ao neurônio i da camada oculta.

Após a aplicação da função de ativação σ sobre o produto pontual de W por X, tem-se a matriz Σ, representada por:

$$\Sigma = \sigma(WX) = \sigma\left(\begin{bmatrix} w_{11} & w_{12} & \cdots & w_{1p} & b_1 \\ w_{21} & w_{22} & \cdots & w_{2p} & b_2 \\ \vdots & \vdots & \vdots & \vdots & \vdots \\ w_{N1} & w_{N2} & \cdots & w_{Np} & b_N \end{bmatrix} \begin{bmatrix} x_{11} & x_{12} & \cdots & x_{1T} \\ x_{21} & x_{22} & \cdots & x_{2T} \\ \vdots & \vdots & \vdots & \vdots \\ x_{p1} & x_{p2} & \cdots & x_{pT} \\ 1 & 1 & \cdots & 1 \end{bmatrix}\right) =$$

$$\begin{bmatrix} \sigma(\mathbf{w_1 x_1} + b_1) & \sigma(\mathbf{w_2 x_1} + b_2) & \cdots & \sigma(\mathbf{w_N x_1} + b_N) \\ \sigma(\mathbf{w_1 x_2} + b_1) & \sigma(\mathbf{w_2 x_2} + b_2) & \cdots & \sigma(\mathbf{w_N x_2} + b_N) \\ \vdots & \vdots & \vdots & \vdots \\ \sigma(\mathbf{w_1 x_T} + b_1) & \sigma(\mathbf{w_2 x_T} + b_2) & \cdots & \sigma(\mathbf{w_N x_T} + b_N) \end{bmatrix} \tag{3.2}$$

Na notação acima, tem-se que $\mathbf{w_i}$ $(i = 1, 2, ..., N)$ representa os pesos do neurônio oculto i, enquanto que b_i é o *bias* associado ao mesmo e $\mathbf{x_j}$, $j = 1, 2, ..., T$ é a amostra j. Outra notação importante aqui é $w_i x_j$, que representa o produto interno entre dois vetores.

Seja então a matriz M dos pesos que conectam a camada oculta com a camada de saída:

$$M = \begin{bmatrix} m_{11} & m_{12} & \cdots & m_{1N} \\ m_{21} & m_{22} & \cdots & m_{2N} \\ \vdots & \vdots & \vdots & \vdots \\ m_{d1} & w_{d2} & \cdots & w_{dN} \end{bmatrix} \tag{3.3}$$

Cada linha i da matriz M está representando os pesos que estão associados ao neurônio de saída s_i. Com isto, o problema de determinar a matriz de pesos M acaba recaindo na solução do sistema linear:

$$M\Sigma = Y. \tag{3.4}$$

O sistema acima pode ser resolvido, por exemplo, com a utilização do método de mínimos quadrados, cuja solução pode ser representada por

$$M = Y\Sigma^t \left(\Sigma\Sigma^t\right)^{-1}. \tag{3.5}$$

Cabem aqui algumas observações sobre os valores numéricos dos pesos que conectam a camada de entrada com a camada oculta da rede. De acordo com (HUANG et al., 2006), estes valores podem ser determinados aleatoriamente. Já em (SÁ JUNIOR; BACKES, 2016), os autores trabalharam com números pseudo aleatórios. Neste trabalho foram utilizados números aleatórios gerados a partir de uma distribuição normal [1] com média zero e desvio padrão um.

As matrizes X e Y também foram alteradas, como será visto em seguida, considerando-se o valor dos pixels da imagem original. Ambas foram normalizadas para valores entre 0 e 1, com a multiplicação dessas matrizes por um fator 1/255.

3.1.1 Pseudo-Inversa Moore-Penrose

A matriz pseudo-inversa de Moore-Penrose (MOORE, 1920; PENROSE, 1955) visa generalizar o conceito clássico de matriz inversa. Por exemplo, dado o sistema linear $Ax = b$, com $A_{n\times n} \in \mathbb{R}^n \times \mathbb{R}^n$, x e b pertencendo ao $\mathbb{R}^n$, se A é não singular,

[1] Foi utilizado o comando do Python numpy.random.normal com seed = 1.

então a solução do sistema linear é única, representada por $x = A^{-1}b$, sendo A^{-1} a matriz inversa de A.

Consideramos um sistema linear

$$Ax = b$$

com $A_{m \times n} \in \mathbb{R}^m \times \mathbb{R}^n$, $m \geqslant n$, $x \in \mathbb{R}^n$ e $b \in \mathbb{R}^m$.

De acordo com o teorema da decomposição em valores singulares (SVD na sigla em inglês) (WATKINS, 2002), para uma matriz A do sistema linear anterior, se $A \neq 0$, então existem matrizes $U \in \mathbb{R}^m \times \mathbb{R}^m$ e $V \in \mathbb{R}^n \times \mathbb{R}^n$ ortogonais tais que:

$$A = UDV^t,$$

em que a matriz D é diagonal com a seguinte forma:

$$D = \begin{bmatrix} d_1 & & & & & & \\ & d_2 & & & & & \\ & & \ddots & & & & \\ & & & d_r & & & \\ & & & & 0 & & \\ & & & & & \ddots & \\ & & & & & & 0 \end{bmatrix} \qquad d_1 \geqslant d_2 \geqslant \cdots \geqslant d_r > 0,$$

$D \in \mathbb{R}^m \times \mathbb{R}^n$ e r é o *posto* da matriz A.

A matriz pseudo-inversa de A é definida então por:

$$A^\dagger = VD^\dagger U^t,$$

com $D^\dagger \in \mathbb{R}^n \times \mathbb{R}^m$:

$$D^\dagger = \begin{bmatrix} 1/d_1 & & & & & & \\ & 1/d_2 & & & & & \\ & & \ddots & & & & \\ & & & 1/d_r & & & \\ & & & & 0 & & \\ & & & & & \ddots & \\ & & & & & & 0 \end{bmatrix}$$

Uma das aplicações mais populares do conceito de pseudo-inversa encontra-se no seguinte resultado (WATKINS, 2002):

Teorema 1. *Considere $A \in \mathbb{R}^m \times \mathbb{R}^n$, $b \in \mathbb{R}^m$ e $x \in \mathbb{R}^n$, se a norma mínima do problema satisfaz*

$$\| b - Ax \|_2 = \min_{w \in \mathbb{R}^n} \| b - Aw \|_2,$$

então $x = A^{\dagger}b$.

Utilizando-se o Teorema 1, uma solução para a Equação 3.4 pode ser expressa por:

$$M = Y\Sigma^{\dagger}. \tag{3.6}$$

3.1.2 Resultados Teóricos: Aproximação Universal

Na sequência, são apresentados alguns resultados teóricos relacionados aos métodos matemáticos envolvidos com o modelo de máquinas de aprendizado extremo (HUANG et al., 2006).

Teorema 2. *Seja uma rede neural "feedforward" de camada oculta única, com T neurônios na camada oculta e uma função de ativação $\sigma : \mathbb{R} \longrightarrow \mathbb{R}$ infinitamente diferenciável em qualquer intervalo, para T amostras distintas arbitrárias (x_i, y_i), com $x_i \in \mathbb{R}^p$ e $y_i \in \mathbb{R}^d$, para qualquer $w_i \in \mathbb{R}^p$ e $b_i \in \mathbb{R}$ escolhidos aleatoriamente de acordo com qualquer distribuição de probabilidade contínua. Então, com probabilidade 1, a matriz Σ, de saída da camada oculta da rede, é inversível e existe β tal que $\| \Sigma\beta - Y \| = 0$.*

A demostração do teorema acima encontra-se em (HUANG et al., 2006). De acordo com este resultado, se o número de neurônios na camada oculta for igual ao número de colunas de X, pode-se encontrar uma matriz Σ inversível que seja capaz de ajustar os valores dos pesos e *bias* a qualquer entrada e saída propostas. Trata-se de um legitimo resultado de “aproximação universal” para ELMs.

Porém, é muito comum que o número de neurônios na camada oculta seja menor do que o necessário para satisfazer o teorema de modo estrito, principalmente por uma questão prática de eficiência computacional. Neste caso, pode-se recorrer ao resultado abaixo.

Teorema 3. *Dados $\varepsilon > 0$ e a função de ativação $\sigma : \mathbb{R} \longrightarrow \mathbb{R}$ infinitamente diferenciável, existe $\tilde{T} \leqslant T$ tal que, para um número T arbitrário de amostras distintas (x_i, y_i), em que $y_i \in \mathbb{R}^p$ e $y_i \in \mathbb{R}^d$, para qualquer w_i e b_i escolhidos aleatoriamente de $\mathbb{R}^p$ e $\mathbb{R}$, respectivamente, de acordo com qualquer distribuição de probabilidade contínua, então, com probabilidade 1, $\| \Sigma_{T\times\tilde{T}}\beta_{\tilde{T}\times d} - Y_{T\times d} \| \leqslant \varepsilon$.*

A prova também encontra-se em (HUANG et al., 2006).

No próximo capítulo, apresenta-se uma explicação detalhada de como o algoritmo de ELM pode ser usado na extração de descritores locais de uma imagem.

4 Base Metodológica

Neste capítulo são descritos os algoritmos que servem como base para o desenvolvimento da metodologia proposta, a saber, a metodologia ELM para a extração de descritores de imagens de textura, tal como apresentado em (SÁ JUNIOR; BACKES, 2016), e os padrões locais binários apresentados em (OJALA et al., 2002).

4.1 Descritores de Textura Baseados em ELM

A utilização do algoritmo ELM, conforme descrito no capítulo 3, para extrair descritores de uma imagem, implicaria em transformá-la em um vetor coluna, removendo assim as relações entre um pixel e sua vizinhança, sendo que em imagens de um modo geral e especialmente as de textura, sabe-se que estes padrões de vizinhança locais apresentam importância crucial na descrição das mesmas.

Diante deste fato, recomenda-se que as Máquinas de Aprendizado Extremo quando empregadas no reconhecimento de imagens de texturas sejam devidamente adaptadas. Um exemplo eficiente de adaptação neste sentido é apresentado em (SÁ JUNIOR; BACKES, 2016), em que as imagens são caracterizadas por meio dos pesos calculados na matriz M descrita no Capítulo 3.

Os valores x_j da matriz X de entrada são extraídos a partir de uma máscara (S) de ordem $k \times k$ que varre a imagem sequencialmente de modo que o centro de S passa por todos os pixels da imagem (possivelmente com algum tratamento especial nas bordas). A representação destas máscaras pode ser observada na Figura 3 para valores de $k = 3, 5, 7$.

O raio do círculo de pixels na máscara é determinado a partir da distância do pixel central para as bordas das janelas apresentadas na Figura 3. No caso, os raios são $\sqrt{2}$, $\sqrt{8}$ e $\sqrt{13}$, respectivamente.

O pixel central de S é adicionado à matriz Y, enquanto que os demais pixels posicionados sobre o círculo destacado na máscara são associados à coluna x_j de X, seguindo uma certa ordem pré-definida (horária ou anti-horária). Como exemplo,

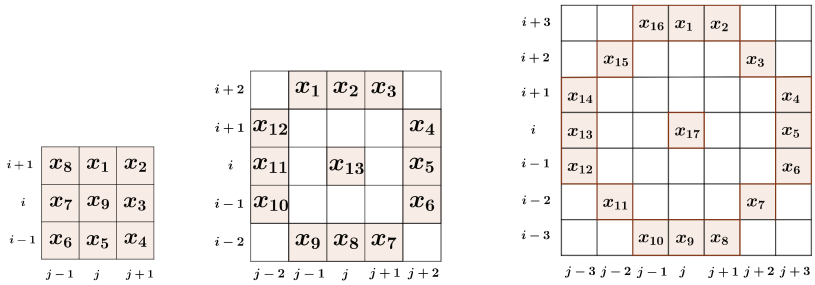


Figura 3 – Submatrizes utilizadas como máscaras no algoritmo de ELM para descritores locais de texturas (SÁ JUNIOR; BACKES, 2016).

para uma máscara 3×3, tem-se:

$$X = \begin{bmatrix} x_1 & : & \cdots & : \\ x_2 & : & \cdots & : \\ : & : & \cdots & : \\ x_8 & : & \cdots & : \end{bmatrix}$$
$$Y = \begin{bmatrix} x_9 & : & \cdots & : \end{bmatrix} \tag{4.1}$$

Os vetores-colunas da matriz X são organizados de modo que para cada submatriz utilizada adotou-se o seguinte esquema: rotacionou-se o vetor $x_j = [x_{1j}, x_{2j} ... x_{i,j}]$, com $i \in \{8, 12, 16\}$, e em cada rotação foi efetuada uma operação de produto interno com os vetores de pesos

$v_3 = [2^{1.0}, 2^{2.0}, 2^{3.0}, 2^{4.0}, 2^{5.0}, 2^{6.0}, 2^{7.0}, 2^{8.0}]$ para $k = 3$.

$v_5 = [2^{1.0}, 2^{1.6}, 2^{2.2}, 2^{2.8}, 2^{3.4}, 2^{4.0}, 2^{4.6}, 2^{5.2}, 2^{5.8}, 2^{6.4}, 2^{7.0}, 2^{7.6}]$ para $k = 5$ e

$v_7 = [2^{1.0}, 2^{1.4}, 2^{1.8}, 2^{2.2}, ..., 2^{7.0}]$ para $k = 7$.

Ao fazer o produto interno entre os vetores resultando em $p_{int} = x_j v_i$, utilizou-se o vetor x_j cujos índices apresentaram o menor valor de p_{int}.

Para cada imagem são gerados então a matriz X e o vetor Y e o algoritmo do ELM pode então ser aplicado conforme já descrito (neste procedimento foi utilizada a solução apresentada na Equação 3.6). Deste modo, o número de elementos do vetor de descritores é igual a N. Caso seja considerado também para compor tal vetor o *bias* para o neurônio de saída, este terá comprimento $N + 1$.

Além das matrizes acima, o algoritmo ELM necessita de uma função de ativação e, neste trabalho, foi utilizada uma do tipo sigmoide.

Para cada máscara apresentada na Figura 3, o algoritmo de ELM é aplicado gerando um vetor de pesos com tamanho igual ao número de neurônios da camada oculta acrescido de mais um elemento de *bias*. A concatenação dos vetores para os três tamanhos de janela resulta no vetor de descritores associado à imagem, o qual é então utilizado para a classificação da textura analisada. Na apresentação dos resultados utilizando esta metodologia, empregou-se a notação ELM para estes descritores.

Este método de extrair descritores de imagens serviu aqui como base para as modificações e estudos desenvolvidos neste trabalho.

4.2 Descritores da imagem em outros espaços

O uso de transformações é algo bastante popular desde os primeiros estudos em análise de texturas. Transformadas como Fourier (PEDRINI; SCHWARTZ, 2008), *wavelets* (PEDRINI; SCHWARTZ, 2008) e Gabor (IDRISSA; ACHEROY, 2002) são bem conhecidas, tanto em análise como em processamento de imagens, e têm sido ferramentas fundamentais na solução de vários problemas na visão computacional. Nos últimos anos, outras transformações locais como os padrões locais binários (OJALA et al., 2002), *textons* (LEUNG; MALIK, 2001) e *dense-SIFT* (LIU et al., 2011a) também ganharam destaque.

4.2.1 Padrões Locais Binários

O método dos padrões locais binários (sigla LBP em inglês) (OJALA et al., 2002), baseia-se em transformar a matriz V da imagem original em uma nova matriz U, em que cada pixel é ajustado de acordo com uma associação com seus pixels vizinhos.

Deste modo, o foco também está na vizinhança de cada pixel da imagem. Neste caso, define-se o parâmetro R como sendo o raio dessa vizinhança e o parâmetro P como o número de pontos amostrados (por interpolação linear) sobre este raio para a construção do código binário.

Na Figura 4 são apresentados exemplos da vizinhança considerada para diferentes valores de R e P. Para valores pré-especificados de R e P, o valor (código

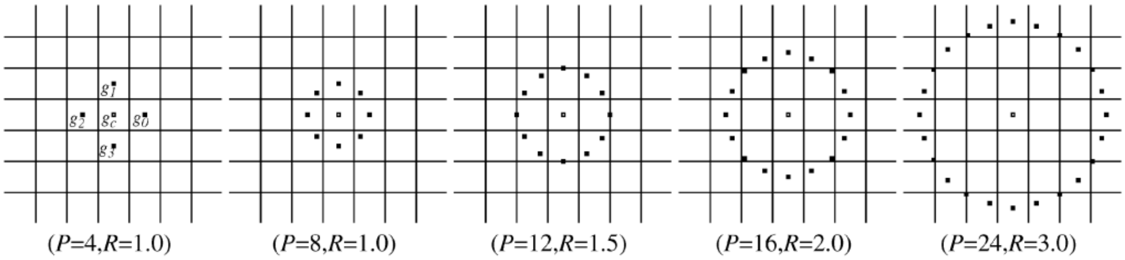


Figura 4 – Vizinhança usada no LBP para diferentes valores de raio R e amostragem P. Fonte (OJALA et al., 2002)

LBP) associado a cada pixel central é calculado da seguinte forma:

$$b_{ij} = \begin{cases} 1, & \text{se} \quad m_{ij} \geqslant m_c \\ 0, & \text{se} \quad m_{ij} < m_c, \end{cases} \tag{4.2}$$

em que m_c é o valor do pixel posicionado sob o centro da vizinhança e m_{ij} são os valores dos P pixels amostrados sobre a circunferência de raio R. Na próxima etapa, seguindo-se uma ordem pré-definida, a sequência de 0's e 1's é usada então para compor um número binário. O pixel u_{ij} recebe então como seu código LBP o valor decimal do número binário associado. A Figura 5 ilustra um exemplo do processo comparativo para $R = 1$.

Um exemplo da conversão de número binário para decimal neste caso pode ser visto na Figura 6. A Figura 7 ilustra os códigos LBP gerados a partir de uma imagem da base de dados UMD.

Os descritores LBP de cada imagem são, por fim, obtidos do histograma da matriz U.

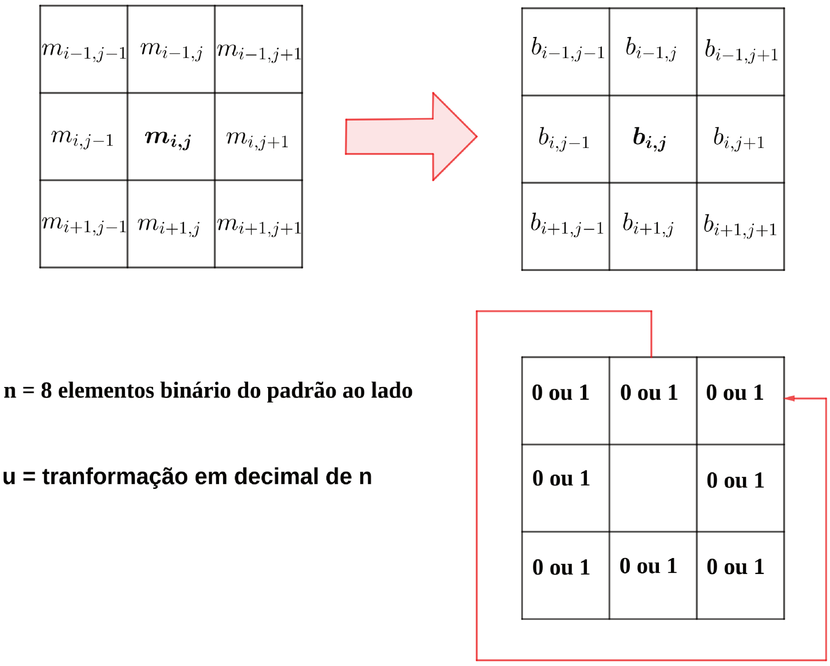


Figura 5 – Código LBP para $R = 1$.

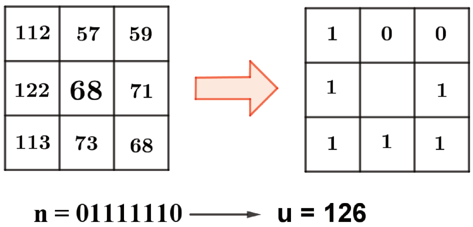


Figura 6 – Exemplo da transformação dos pixels em números binários e posterior conversão para número decimal (código LBP).

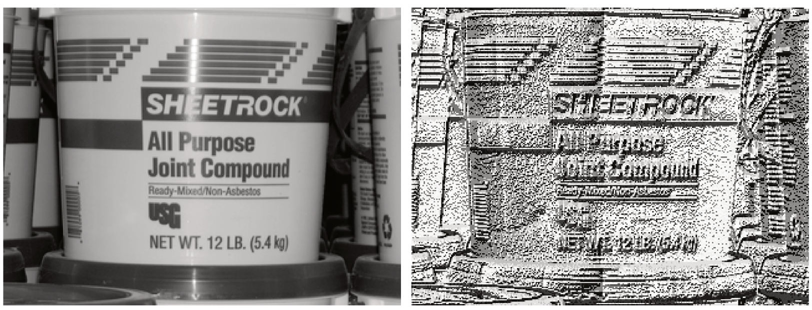


Figura 7 – À esquerda, imagem original, à direita, imagem após substituição de cada pixel por seu código LBP (transformada LBP).

5 Metodologia Proposta

Este estudo trabalha com cinco abordagens para a extração de descritores de texturas usando ELMs: descritor ELM aplicado sobre padrões locais binários (LBP), transformada ELM, descritores de ELM usando LBP apenas na saída da rede, descritores ELM morfológicos e regressão *ridge*.

5.1 ELM sobre a transformada LBP

Esta abordagem explora o uso de transformações/representações locais como entrada do modelo de geração de descritores locais por ELM. Sabe-se que tais transformações oferecem novas perspectivas sobre a textura que a imagem original não consegue representar. Um exemplo bem ilustrativo deste fato é que o simples histograma de uma imagem de textura não consegue discriminar dois materiais diferentes, porém, se este mesmo histograma for extraído de um mapa de padrões binários (LBP), estes podem ser discriminados por medidas simples de diferença entre histogramas (distância *chi*-quadrado, por exemplo) (Ojala et al., 2002).

O método de extração dos descritores ELM a partir de uma transformada LBP é esquematizado na Figura 8:

Figura 8 – ELM aplicado sobre a transformada LBP.

Após a entrada da imagem no algoritmo, aplica-se uma transformada LBP, como descrito na Seção 4.2.1, isto é, cada pixel na imagem original é substituído pelo código LBP (decimal) correspondente. Neste trabalho, usou-se $R = 1$ e $P = 8$ para essa transformação. Na sequência, o algoritmo de ELM é aplicado gerando assim o vetor de descritores da imagem. Para o número de neurônios na camada oculta da rede do modelo ELM foram testados os valores $N = \{9, 19, 29, 39, 49, 59\}$.

Nessa abordagem, existe uma transformação em todos os pixels da imagem, como ilustrado na Figura 7. Cada elemento na transformada LBP possui uma relação associada a uma determinada vizinhança dos pixels correspondentes na imagem original, causando assim mudança nos valores dos dados das matrizes X e Y quando comparados com o ELM original apresentado na Seção 4.1.

O uso desta nova configuração torna possível a obtenção de um descritor de textura com outra perspectiva sobre a imagem original, contribuindo assim para o enriquecimento do vetor de características e potencial aumento na acurácia do processo de classificação. Na apresentação dos resultados, foi adotada a notação ELM_{LBP} para esta solução proposta.

5.2 ELM com LBP na saída

A segunda abordagem proposta baseia-se também em combinar o algoritmo ELM com LBP (ELM+LBP), porém agora usando códigos LBP apenas na saída da rede ELM, mantendo a entrada a partir dos pixels da imagem original. Deste modo, a configuração do vetor Y no algoritmo ELM foi alterada e um esquema do modelo pode ser observado na Figura 9.

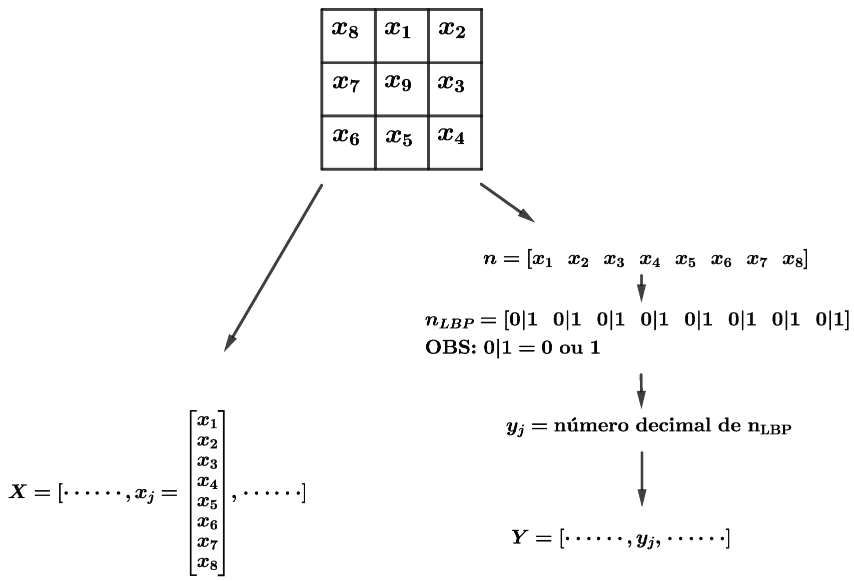


Figura 9 – Algoritmo ELM com códigos LBP na saída da rede.

Assim como na metodologia originalmente proposta em (SÁ JUNIOR; BACKES, 2016), a matriz X é formada diretamente a partir dos valores dos pixels da imagem de textura analisada. Porém, o vetor Y, em vez do valor do pixel central da submatriz, agora recebe o código LBP decimal correspondente ao pixel cujos vizinhos são armazenados na coluna x_j da matriz de entrada.

Neste procedimento, portanto, temos a combinação do LBP com o ELM em uma só ponta do processo, ou seja, na saída, diferentemente da metodologia na Seção 4.2 em que ambas as partes da rede são alteradas. Na seção de resultados, esta abordagem é representada pela notação $ELM + LBP$.

5.3 Transformação ELM

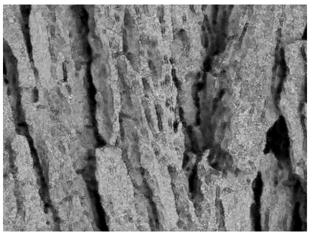
Quando o vetor de pesos do modelo ELM de uma imagem é calculado, todos os pixels da imagem contribuem para a construção deste vetor, já que para este cálculo o algoritmo percorre a imagem com uma submatriz e organiza os dados em amostras de entrada (pixels vizinhos) e saída (pixel central). Essa ideia de explorar a vizinhança dos pixels e a forma como esta se relaciona com o pixel central é amplamente explorada em transformações de imagens na literatura. E esta é a motivação para a proposta aqui do que chamamos de “transformação ELM”.

Esse algoritmo funciona na seguinte sequência: primeiramente é aplicado o ELM tal como descrito na Seção 4.1 sobre a imagem original. Dessa forma, calcula-se o vetor de pesos (v_p). Na segunda etapa, o ELM é usado como preditor, ou seja, aplicamos o algoritmo tendo como entrada as vizinhanças da imagem original, mas utilizando os valores do vetor v_p entre a camada oculta e de saída. Por fim, os valores da saída da rede são usados como valor do pixel correspondente na imagem transformada, gerando dessa forma uma reconstrução da imagem original em que cada pixel recebe influência direta de toda a sua estrutura de vizinhança.

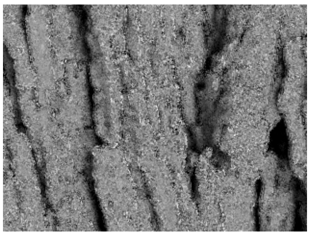
Um exemplo visual desta transformação de imagem é apresentado na Figura 10. Nela foi variado o raio do ELM e o número de neurônios na camada oculta. Quanto maior o raio, mais se percebe um embaçamento (*blurring*) na imagem original após a transformada ELM.

Quando aplicado o ELM sobre a imagem, pela própria construção do algoritmo, os pixels das bordas não são armazenados na matriz Y. Com isto, a nova imagem após a transformação ELM fica menor do que a original.

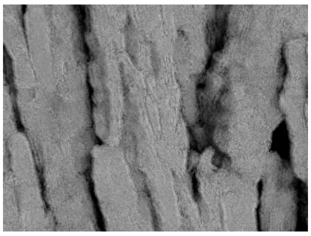
(a) Imagem original.



(b) r=$\sqrt{2}$ e 3 N.

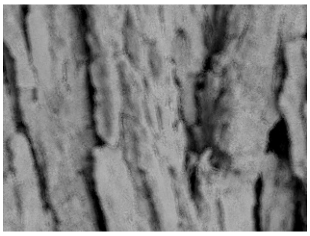


(c) r=$\sqrt{8}$ e 3 N.

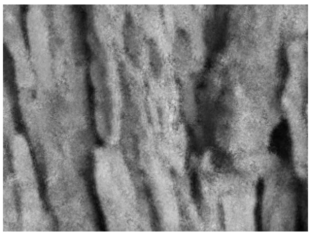


(d) r=$\sqrt{13}$ e 3 N.

(e) r=$\sqrt{2}$ e 6 N.



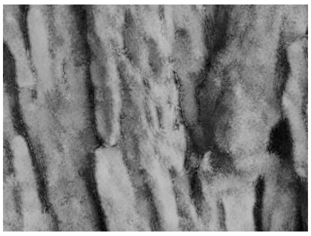
(f) r=$\sqrt{8}$ e 6 N.



(g) r=$\sqrt{13}$ e 6 N.

(h) r=$\sqrt{2}$ e 9 N.

(i) r=$\sqrt{8}$ e 9 N.



(j) r=$\sqrt{13}$ e 9 N.

Figura 10 – Aplicação da Transformada ELM em uma imagem, com variação no raio r e no números de neurônios N da camada oculta do ELM.

Conforme aumenta-se o número de neurônios da camada oculta, a imagem transformada tende ao seu estado original, fato este esperado pelo menor erro na aplicação do método de mínimos quadrados no ELM. Já com número reduzido de neurônios, especialmente para $N = 3$, a degradação nos pixels se torna visível. O aspecto visual se torna similar ao de uma imagem altamente compactada por algum formato com perda, como *JPG* por exemplo.

A partir da imagem transformada, pode-se aplicar qualquer operação de processamento ou análise de imagens. Aqui, em particular, usa-se o algoritmo LBP para obter-se o vetor de características para a posterior classificação dessas imagens. Um esquema geral deste modelo está representado na Figura 11. Nas tabelas com os resultados da classificação, foi utilizada a notação LBP_{ELM} para esse método.

$\boxed{\textbf{Imagem}} \rightarrow \boxed{\textbf{Transformação ELM}} \rightarrow \boxed{\textbf{LBP}} \rightarrow \boxed{\textbf{Vetor de descritores}}$

Figura 11 – Transformada ELM.

Para a aplicação desta nova abordagem, é necessário que se definam os parâmetros tanto do ELM quanto do LBP. Nos testes realizados nas seções de resultados, chegou-se empiricamente a configurações do ELM com máscara 3×3 e raio $\sqrt{2}$ em todos os casos. Para o número de neurônios na camada oculta foram testados $N = \{3, 6, 9, 19, 29, 39, 49, 59\}$. Em todos os casos, usou-se uma função de ativação sigmoide. Para o LBP foi usado $R = 1$ e $P = 8$.

5.4 ELM Morfológico

Em (SUSSNER; CAMPIOTTI, 2020), os autores trabalham com a adaptação de uma rede neural morfológica para o ELM clássico, tal como proposto em (HUANG et al., 2006). O método é chamado de perceptron morfológico/linear híbrido (sigla HMLP, do inglês *hybrid morphological/linear perceptron*). Entre as vantagens das máquinas de aprendizado extremo neste tipo de metodologia, destaca-se o fato de que o algoritmo baseado em mínimos quadrados elimina o problema da ausência de diferenciabilidade nas redes morfológicas. O modelo ELM torna-se assim uma ferramenta ao mesmo tempo simples e poderosa para implementações de arquiteturas morfológicas, permitindo assim tanto seu estudo do ponto de vista teórico quanto a exploração desse tipo de abordagem na prática.

Destacaremos alguns conceitos matemáticos explorados nessa abordagem. Definindo-se um *poset* $\mathbb{L}$, uma rede completa $\mathbb{M}$ e $X \subseteq \mathbb{L}$, como descritos em (SUSSNER; CAMPIOTTI, 2020), uma *erosão* é um operador $\epsilon : \mathbb{L} \rightarrow \mathbb{M}$ que comuta com a operação mínimo:

$$\epsilon(\bigwedge X) = \bigwedge \epsilon(X)$$

Já a *dilatação* $\delta : \mathbb{L} \rightarrow \mathbb{M}$ comuta com a operação máximo:

$$\delta(\bigvee X) = \bigvee \delta(X)$$

Anti-erosão e *Anti-dilatação* são operadores $\bar{\epsilon}, \bar{\delta} : \mathbb{L} \rightarrow \mathbb{M}$, respectivamente definidos por:

$$\bar{\epsilon}(\bigwedge X) = \bigvee \bar{\epsilon}(X)$$

$$\bar{\delta}(\bigvee X) = \bigwedge \bar{\delta}(X)$$

O simbolo $\wedge$ representa o *mínimo* enquanto $\vee$, o *máximo.*

Um exemplo da aplicação de erosões e dilatações morfológicas em processamento de imagens pode ser observado na Figura 12. Neste trabalho, a metodologia

(a) Imagem original. (b) Erosão. (c) Dilatação.

Figura 12 – Operações morfológicas em uma imagem binária. Fonte:(BRADSKI, 2000).

descrita em (SUSSNER; CAMPIOTTI, 2020) é generalizada para a extração de descritores locais de uma imagem de textura. Um esquema geral de ELM morfológico é apresentado na Figura 13. Nesta figura, o vetor x representa uma amostra de entrada ($x \in \mathbb{R}^p$), M é o número de neurônios morfológicos e L indica o número de neurônios lineares (clássicos). Deste modo, o número total de neurônios da camada oculta é representado por $N = M + L$.

Nessa abordagem, além da matriz de pesos W, também são necessárias mais duas matrizes C e D, ambas pertencendo a $\mathbb{R}^p \times \mathbb{R}^M$, e determinadas aleatoriamente de modo que $-c_{ij} \leqslant d_{ij}$.

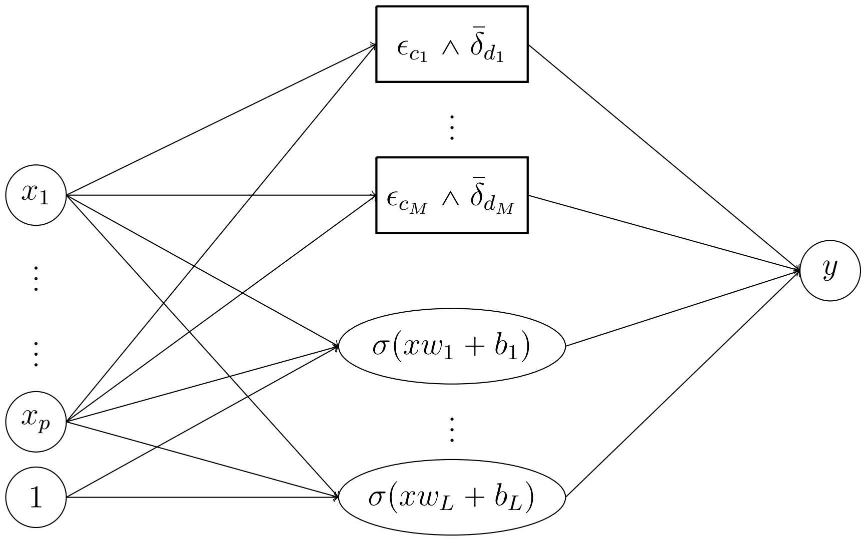


Figura 13 – Representação do ELM hibrido/linear, adaptado de (SUSSNER; CAMPIOTTI, 2020)

Nos neurônios morfológicos, são efetuadas operações de erosão e anti-dilatação (SUSSNER; CAMPIOTTI, 2020), definidas como:

$$\epsilon_c(x) \wedge \bar{\delta}_d(x) = \bigwedge_{i=1}^{p}(x_i + c_i) \wedge \bigwedge_{i=1}^{p}(d_i - x_i) = \bigwedge_{i=1}^{p}[(x_i + c_i) \wedge (d_i - x_i)] \tag{5.1}$$

Aqui o símbolo $\wedge$ representa o mínimo, x é um vetor-coluna da matriz X de amostras, enquanto que c e d são vetores-colunas das matrizes C e D, respectivamente.

Neste procedimento, a matriz Σ da Equação 3.2 assume a seguinte forma:

$$\Sigma^t = \begin{bmatrix} \epsilon_{c_1}(x_1) \wedge \bar{\delta}_{d_1}(x_1) & \cdots & \epsilon_{c_M}(x_1) \wedge \bar{\delta}_{d_M}(x_1) & \sigma(w_1x_1 + b_1) & \cdots & \sigma(w_Lx_1 + b_L) \\ \epsilon_{c_1}(x_2) \wedge \bar{\delta}_{d_1}(x_2) & \cdots & \epsilon_{c_M}(x_2) \wedge \bar{\delta}_{d_M}(x_2) & \sigma(w_1x_2 + b_1) & \cdots & \sigma(w_Lx_2 + b_L) \\ \vdots & \vdots & \vdots & \vdots & \vdots & \vdots \\ \epsilon_{c_1}(x_T) \wedge \bar{\delta}_{d_1}(x_T) & \cdots & \epsilon_{c_M}(x_T) \wedge \bar{\delta}_{d_M}(x_T) & \sigma(w_1x_T + b_1) & \cdots & \sigma(w_Lx_T + b_L) \end{bmatrix} \tag{5.2}$$

Deste modo, chega-se a um sistema idêntico ao da Equação 3.4, o qual pode ser resolvido pelo mesmo procedimento do ELM padrão.

Nos testes de extração de descritores de imagens deste trabalho, foram utilizadas duas variantes para este método. Na primeira, foram extraídos os vetores de características com nós tanto lineares quanto morfológicos na camada oculta. Por exemplo, para um total de 9 neurônios foram usados 5 lineares e 4 morfológicos, para 19 neurônios podemos ter 10 lineares e 9 morfológicos, e assim sucessivamente. Para esta abordagem foi utilizada a notação ELM_{MM}. Na segunda alternativa, foram utilizados apenas nós morfológicos na camada oculta, solução essa aqui representada por ELM_{morf}.

5.5 Uso de Variantes Mais Elaboradas de ELMs: Regressão *Ridge*

Para calcular a solução do sistema linear $M\Sigma = Y$, na Equação 3.5 é necessário calcular uma matriz inversa $(\Sigma\Sigma^t)^{-1}$. Caso Σ apresente multicolinearidade, isto é, colunas linearmente dependentes, ou pelo menos próximas de uma dependência linear, $\Sigma\Sigma^t$ torna-se quase singular e os paramentos de M podem ter variância que não seja a mínima.

Uma das alternativas propostas para este tipo de problema, (CALVETTI et al., 2000; TIKHONOV, 1963) baseia-se em adicionar um fator regularizador de modo a transformar o problema de inversão original no seguinte:

$$M(\Sigma\Sigma^t + \gamma I) = Y\Sigma^t$$

Deste modo, temos a solução:

$$M = Y\Sigma^t(\Sigma\Sigma^t + \gamma I)^{-1},$$

com $0 < \gamma$ e sendo I a matriz identidade.

A partir do bom desempenho alcançado pelo algoritmo de ELMs em textura, já em sua versão original, a literatura tem mostrado evoluções que têm contribuído sobremaneira para uma melhor performance (BACKES et al., 2019; RIBAS et al., 2020), especialmente em problemas mais desafiadores como o reconhecimento de imagens de texturas aqui tratado.

Um exemplo nesta linha é o ELM regularizado, em que os pesos são obtidos por um processo de regressão *ridge*, com parâmetro regularizador $\gamma > 0$, como descrito no Capítulo 6. Além de permitir a obtenção de descritores com menor variação e maior robustez à variação local e ruídos, o estudo destes descritores para diferentes valores de γ é interessante, tanto do ponto de vista teórico quanto prático. Os descritores para uma família de γ podem, por exemplo, ser concatenados, para assim somarem forças e produzirem um descritor local ainda mais preciso e robusto. Os resultados encontrados com esta abordagem são aqui representados por ELM_{ridge}.

Vale destacar que a utilização de regularização para resolver a Equação 3.4 já tem sido explorada na literatura. A novidade neste trabalho é uma análise do quanto diferentes valores de γ podem influenciar na taxa de acerto.

5.6 Aplicações

A metodologia proposta foi aplicada a um problema prático de grande interesse na área médica, que é a identificação da presença ou não de câncer de pulmão a partir de exames citológicos, bem como a identificação do tipo de câncer presente em cada caso.

O câncer de pulmão está entre os mais letais em todo o mundo, sendo o câncer que mais mata entre homens e o segundo entre mulheres (INCA, 2020) e está também entre os mais agressivos. Particularmente, no caso de um dos tipos deste câncer, o neuroendócrino de pequenas células, 60% das vezes em que este é diagnosticado por técnicas convencionais, já se encontra na fase de metástase, evoluindo muito rapidamente. O diagnóstico e identificação correta do tipo de câncer neste caso é fundamental e o auxílio de ferramentas de visão computacional ao médico patologista tem se mostrado muito promissor.

6 Análise Assintótica

Não há como negar que as redes neurais modernas têm alcançado resultados extraordinários nas mais diversas tarefas. Porém, o que se vê na imensa maioria das vezes é que estas redes “aprendem” a partir de uma gigantesca quantidade de dados para treinamento e de um sistema computacional muito poderoso. Não é exagero dizer que o aprendizado em questão se dá por “força bruta”. Ao mesmo tempo, são raros os estudos que buscam entender melhor os mecanismos de funcionamento de uma rede neural a partir de ferramentas formais de análise matemática. Isto se dá em grande parte pela complexidade destes modelos, contendo muitas vezes dezenas ou até mesmo centenas de camadas. Cabe ressaltar, porém, que estudos deste tipo são imprescindíveis para um avanço melhor direcionado da área.

No caso específico dos ELMs, como já foi discutido, o fato de sua arquitetura típica recair em um sistema cuja solução em tempo de treinamento é exato e direto traz grande vantagem do ponto de vista computacional. Porém, esta característica também traz uma outra implicação bastante interessante, que é a possibilidade de se estabelecer um modelo matemático que responda, ao menos assintoticamente e em termos de distribuições estatísticas, algumas perguntas importantes relativas a problemas deste tipo.

Uma das ferramentas que tem se mostrado mais adequada para uma análise formal deste tipo de modelo é a teoria das matrizes aleatórias. Como já foi descrito, o estado dos neurônios intermediários é definido pelos dados de entrada, assim como também pelos pesos aleatórios, de modo que os próprios estados acabam sendo variáveis aleatórias também. É fato também que ELMs possuem relação com redes neurais aleatórias (GELENBE, 1989), que são tipicamente estudadas na matemática por matrizes aleatórias. Por fim, vale lembrar que dados em geral, estejam eles presentes no conjunto de treinamento ou de teste, são usualmente modelados como variáveis aleatórias que seguem uma determinada distribuição (frequentemente uma Gaussiana).

O estudo deste capítulo analisa os ELMs teoricamente, tratando-os como um modelo de regressão linear "ridge" (LOUART et al., 2018) (regularizado). A entrada do ELM será composta por $X = [\mathbf{x_1}, \cdots, \mathbf{x_T}] \in \mathbb{R}^{p \times T}$ e a saída por $Y = [\mathbf{y_1}, \cdots, \mathbf{y_T}] \in \mathbb{R}^{d \times T}$. Os pesos entre a entrada e a camada escondida são valores aleatórios representados em uma matriz $W \in \mathbb{R}^{n \times p}$. A função de ativação $\sigma : \mathbb{R} \to \mathbb{R}$ é não linear é aplicada ponto a ponto sobre matrizes e vetores. O mapa de características na camada intermediária é dado então por $\Sigma \equiv \sigma(WX) \in \mathbb{R}^{n \times T}$.

Como um problema clássico de regressão, para cada entrada $\mathbf{x} \in \mathbb{R}^p$, a saída $\mathbf{z} \in \mathbb{R}^d$ da rede é dada por

$$\mathbf{z} = \beta^\top \sigma(W\mathbf{x}),$$

em que $\beta \in \mathbb{R}^{n \times d}$ é uma matriz que será encontrada no treinamento. No treinamento, com base em X e Y, β é obtida por

$$\arg\min_\beta \frac{1}{T} \sum_{i=1}^{T} \|\mathbf{z_i} - \mathbf{y_i}\|^2 + \gamma \|\beta\|_F^2,$$

em que $\gamma > 0$ é um fator regularizador. Resolvendo-se este problema de regressão *ridge* obtém-se

$$\beta = \frac{1}{T} \Sigma \left(\frac{1}{T} \Sigma^\top \Sigma + \gamma I_T \right)^{-1} Y^\top, \tag{6.1}$$

em que I_T é a matriz identidade $T \times T$. Uma parte fundamental da Equação (6.1) é aquela que em seguida será definida como matriz Q:

$$Q \equiv \left(\frac{1}{T} \Sigma^\top \Sigma + \gamma I_T \right)^{-1}.$$

Com esta definição, o erro quadrático médio no treinamento pode então ser elegantemente formulado como

$$E_{treino} = \frac{1}{T} \left\| Y^\top - \Sigma^\top \beta \right\|_F^2 = \frac{\gamma^2}{T} \operatorname{tr}(Y^\top Y Q^2).$$

Já na fase de teste do ELM, dado um conjunto de dados $\hat{X} \in \mathbb{R}^{p \times \hat{T}}$, o erro médio quadrático de teste é

$$E_{teste} = \frac{1}{T} \left\| \hat{Y}^\top - \hat{\Sigma}^\top \beta \right\|_F^2,$$

em que $\hat{\Sigma} \equiv \sigma(W\hat{X})$.

Para viabilizar a análise seguinte, algumas suposições serão necessárias. Primeiramente, a matriz W deve ser *subgaussiana*, isto é, pode-se escrever $W = \varphi(\tilde{W})$, em que as entradas de $\tilde{W}$ são independentes e geradas por uma distribuição normal padronizada e $\varphi(\cdot)$ é Lipschitz com parâmetro λ_φ. A função σ também deve ser contínua e Lipschitz com parâmetro λ_σ. Por fim, no limite quando $n \to \infty$, γ, λ_φ, λ_σ e d devem permanecer constantes e

$$0 < \liminf_n \min\{p/n, T/n\} \leqslant \limsup_n \max\{p/n, T/n\} < \infty,$$

além de

$$\limsup_n \|X\| < \infty, \qquad \limsup_n \max_{ij} |Y_{ij}| < \infty.$$

Bastando apenas as duas primeiras condições como válidas (isto é sem $n \to \infty$), o seguinte lema pode ser estabelecido.

Lema 1. *Seja Φ uma matriz $T \times T$ definida por*

$$\Phi = \mathrm{E}\left[\sigma\left(w^\top X\right)^\top \sigma\left(w^\top X\right)\right],$$

em que $w \sim \mathcal{N}_\varphi(0, I_p)$. Seja ainda a matriz $A \in \mathbb{R}^{T\times T}$ tal que $\|A\| \leqslant 1$ e defina-se o vetor $\sigma \equiv \sigma\left(w^t X\right)^t \in \mathbb{R}^T$. Então

$$P\left(\left|\frac{1}{T}\sigma^\top A\sigma - \frac{1}{T}\mathrm{tr}\Phi A\right| > t\right) \leqslant C\mathrm{e}^{-\frac{cT}{\|X\|^2\lambda_\varphi^2\lambda_\sigma^2}\min\left(\frac{t^2}{t_0^2}, t\right)},$$

em que P denota probabilidade e $t_0 \equiv |\sigma(0)| + \lambda_\varphi\lambda_\sigma\|X\|\sqrt{\frac{p}{T}}$ e sendo c e C constantes positivas independentes. Se a terceira condição também valer:

$$P\left(\left|\frac{1}{T}\sigma^\top A\sigma - \frac{1}{T}\mathrm{tr}\Phi A\right| > t\right) \leqslant C\mathrm{e}^{-cn\min(t,t^2)}.$$

Com este lema em mãos, pode-se estabelecer o seguinte teorema, no caso em que as três condições sejam válidas.

Teorema 4. *Defina-se a matriz $\overline{Q}$ por*

$$\overline{Q} \equiv \left(\frac{n}{T}\frac{\Phi}{1+\delta} + \gamma I_T\right)^{-1},$$

em que δ é a solução (única) de $\delta = \frac{1}{T}\mathrm{tr}(\Phi\overline{Q})$. Então, para todo $\epsilon > 0$, existe $c > 0$ tal que o valor esperado de Q satisfaz

$$\|E[Q] - \overline{Q}\| \leqslant cn^{-\frac{1}{n}+\epsilon}.$$

Por fim, definindo-se a matriz

$$\Psi = \frac{n}{T}\frac{\Phi}{1+\delta},$$

chega-se ao teorema mais importante desta análise do comportamento assintótico do erro quadrático médio de treinamento.

Teorema 5. *Assumindo-se a validade das três condições descritas anteriormente e definindo-se*

$$E_{treino} = \frac{1}{T}\left\|Y^\top - \Sigma^\top\beta\right\|_F^2 = \frac{\gamma^2}{T}\mathrm{tr}(Y^\top Y Q^2)$$

$$\overline{E}_{treino} = \frac{\gamma^2}{T}\mathrm{tr}Y^\top Y\overline{Q}\left[\frac{\frac{1}{n}\mathrm{tr}\overline{Q}^2}{1-\frac{1}{n}\mathrm{tr}(\Psi\overline{Q})^2}\Psi + I_T\right]\overline{Q},$$

então

$$n^{\frac{1}{n}-\epsilon}(E_{treino} - \overline{E}_{treino}) \to 0$$

quase certamente.

Toda essa análise pode ser estendida também para a fase de testes. Porém para o propósito deste estudo, em que usam-se ELMs exclusivamente para a extração de descritores locais, o resultado na fase de treinamento é mais útil.

Realizando uma análise dos resultados anteriores, supondo n suficientemente grande, pelo Teorema 2, temos que a parcela E_{treino} converge assintoticamente para 0. E o fator $n^{\frac{1}{n}-\epsilon}$ converge para 1 no Teorema 5.

Para a parcela $\overline{E}_{treino}$, $\frac{1}{n}$ tende a 0 e, com isto, $\frac{\frac{1}{n}\mathrm{tr}\overline{Q}^2}{1-\frac{1}{n}\mathrm{tr}(\Psi\overline{Q})^2} \to 0$ e $\overline{E}_{treino}$ satisfaz:

$$\overline{E}_{treino} = \frac{\gamma^2}{T}\mathrm{tr}Y^\top Y\overline{QQ},$$

Pelo Teorema 4, o valor esperado da matriz Q tende a $\overline{Q}$, logo:

$$\overline{E}_{treino} = \frac{\gamma^2}{T}\mathrm{tr}Y^\top Y\overline{QQ} \to \frac{\gamma^2}{T}\mathrm{tr}Y^\top YQQ = E_{treino}$$

Portanto analisar E_{treino} e $\overline{E}_{treino}$ acaba sendo equivalente.

Nos trabalhos (LOUART et al., 2018) e (LOUART; COUILLET, 2017), os autores utilizam a teoria apresentada anteriormente para analisar parâmetros da rede neural, como o valor de γ, números de neurônios da camada oculta e função de ativação. Nas simulações apresentadas pelos autores, são destacadas a variação no valor de γ e diferentes funções de ativação para uma parte dos dados da base MNIST (LECUN et al., 1998).

Neste trabalho de mestrado, foi testada uma versão da teoria da análise assintótica adaptada para o caso da utilização de ELM para a extração de descritores.

Com isto, os dados de entrada para a simulação foram referentes à extração de vetores utilizando uma máscara 3×3 em uma imagem de textura. Nos resultados a seguir, usou-se um exemplo de imagem da base de dados UMD (XU et al., 2009b), devido à sua alta resolução. A matriz X do modelo foi composta pelos valores dos pixels vizinhos em relação ao central, que por sua vez formou o vetor Y.

O primeiro objetivo foi analisar o erro de aproximação para diferentes valores de γ. Uma visualização gráfica desta simulação pode ser observada na Figura 14.

A variação de γ foi de 10^{-4} a 100, com os menores valores apresentando o menor erro. Com isto, foram adotados valores entre 10^{-4} a 1 para o algoritmo ELM com regressão *ridge* neste trabalho. Normalmente, para a extração de descritores são adotados valores de $\gamma \leqslant 10^{-1}$, por exemplo em (SÁ JUNIOR et al., 2019) utilizaram o valor de 10^{-3}, além destes valores próximos de zero, também foi utilizado os valores 0.5 e 1, para analisar o comportamento da taxa de acerto nas classificações de imagens.

Em uma segunda análise, foi fixado um valor para γ e variou-se o número de neurônios da camada oculta, chegando-se assim aos valores apresentados na Figura 15.

A variação do erro foi pequena, mas a partir de 60 neurônios na camada oculta do ELM a diferença foi menor que 0,005. Estes resultados possibilitam conclusões semelhantes aos encontrados na seção de resultados e servem como indicativo de que o uso de um número muito alto de neurônios na camada oculta aumenta o custo computacional sem acarretar qualquer ganho representativo na descrição da imagem.

A análise tratada nesta seção serviu então como embasamento teórico tanto para o intervalo de valores de γ considerados na regressão *ridge* quanto para os valores mais adequados para o número de neurônios na camada oculta nos experimentos realizados nos capítulos seguintes.

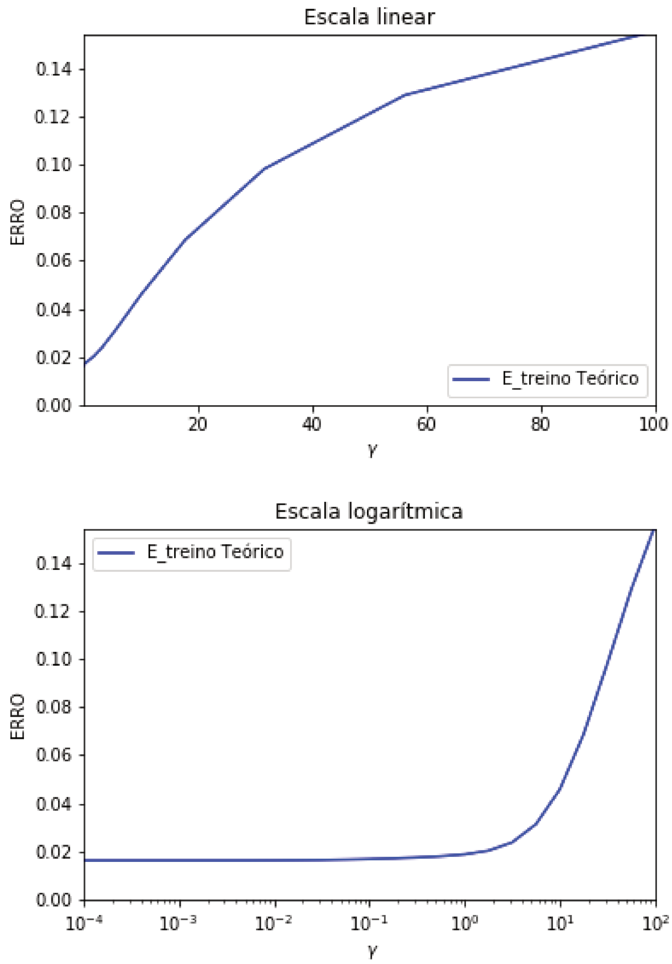


Figura 14 – Curvas de $\overline{E}_{treino}$ para diferentes valores de γ no modelo ELM para descritores de texturas.

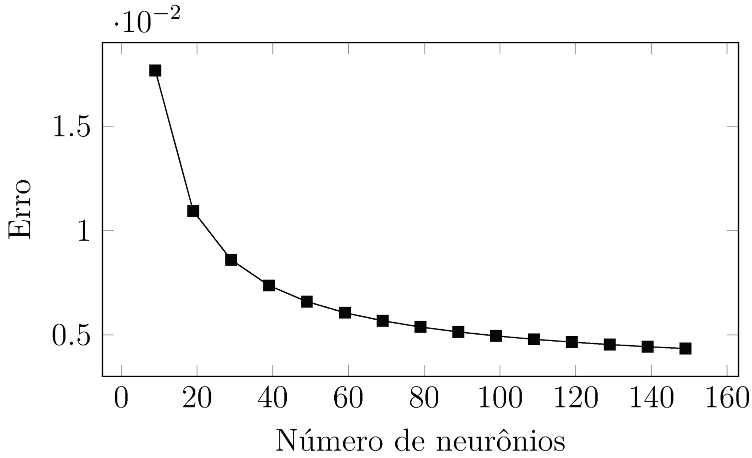


Figura 15 – Variação no número de neurônios da camada oculta do ELM com $\gamma = 10^{-4}$.

7 Experimentos em Bases de *Benchmark*

Para a reprodução dos métodos computacionais e dos resultados discutidos neste estudo foi utilizada a linguagem de programação Python, em sua versão 3.7.4, com o editor Jupyter Notebook, em um computador com sistema operacional Ubuntu 18.04.2 LTS, memória RAM 16 GB, processador Intel® Core™ i5-7500 CPU @ 3.40 GHz $\times$ 4 (sistema e arquitetura de 64 bits).

Os métodos descritos anteriormente servem para extrair um vetor de características de cada imagem. A partir deste conjunto de vetores, faz-se necessário então o uso de um classificador para avaliar o desempenho da metodologia tanto em bases de *benchmark* quanto em aplicações do mundo real.

Neste trabalho foram verificados três classificadores: Análise Discriminante Linear (LDA da sigla em inglês), Máquina de Vetores de Suporte (SVM da sigla em inglês) e Florestas Aleatórias (FAL).

Para fins de cálculo da porcentagem de acerto, usou-se um "limiar duro", isto é, conta-se como "acerto" sempre que a classe correta obteve a maior probabilidade no classificador, mas o valor exato desta probabilidade em si não é considerado. Além desta maneira de avaliar, também foram utilizadas matrizes de confusão, com o objetivo de analisar o desempenho do classificador em cada classe.

Nas próximas seções são apresentados os resultados da classificação de várias bases de imagens. Os resultados são apresentados para diferentes números de neurônios (N) da camada oculta do algoritmo ELM, o que é denotado nas tabelas como $(ELM)_N$, com $N \in \mathbb{N}$. Já a representação da metodologia que aplica o ELM nas imagens após a aplicação da transformada LBP tem a notação $(ELM_{LBP})_N$. O algoritmo ELM usando códigos LBP na saída foi denotado como $(ELM + LBP)_N$, enquanto que o ELM com regressão *ridge* foi denotado como $(ELM_{ridge})_N$. O ELM morfológico foi representado por $(ELM_{MM})_N$ para nós lineares mesclados com morfológicos na camada oculta e $(ELM_{morf})_N$ para apenas neurônios morfológicos.

Também foi explorado o algoritmo de padrões locais binários, que utiliza esta simbologia LBP nas tabelas de classificação. Para a utilização do LBP após a transformada ELM, foi utilizada a notação LBP_{ELM}.

Na sequência, foram concatenados os descritores dos métodos anteriores para fazer a classificação das imagens. O sinal de “+” indica a concatenação de descritores distintos.

Neste trabalho, os algoritmos foram utilizados em três bases de imagens disponibilizadas na literatura: UMD (XU et al., 2009b), UIUC (LAZEBNIK et al., 2005) e KTHTIPS-2b (HAYMAN et al., 2004). As taxas de acerto na classificação foram comparadas com outros trabalhos na literatura que adotaram a mesma versão da base bem como protocolos de treinamento e teste semelhantes.

7.1 Bases de *Benchmark*

Neste capítulo são apresentadas as bases de *benchmark* UMD, UIUC e KTHTIPS-2b, as quais foram avaliadas usando-se as metodologias propostas para a extração de descritores e classificação de imagens.

7.1.1 Texturas UMD

A base de imagens de texturas UMD (XU et al., 2009b) contém um conjunto de 1000 imagens, em tons de cinza, cada uma originalmente com tamanho 1280×960.

A base contém um total de 25 classes com 40 imagens cada e essas imagens apresentam variação na aproximação da câmera com os objetos (escala) e não possuem controle nas condições de luminosidade. As imagens são de objetos diferentes como prateleiras, baldes, frutas, texturas de piso, entre outros. Um exemplo de cada classe pode ser visto na Figura 16.

Devido ao custo computacional envolvido, o tamanho das imagens neste trabalho foi reduzido para 320×240, seguindo o padrão dos trabalhos da literatura usados para comparação das taxas de acertos.

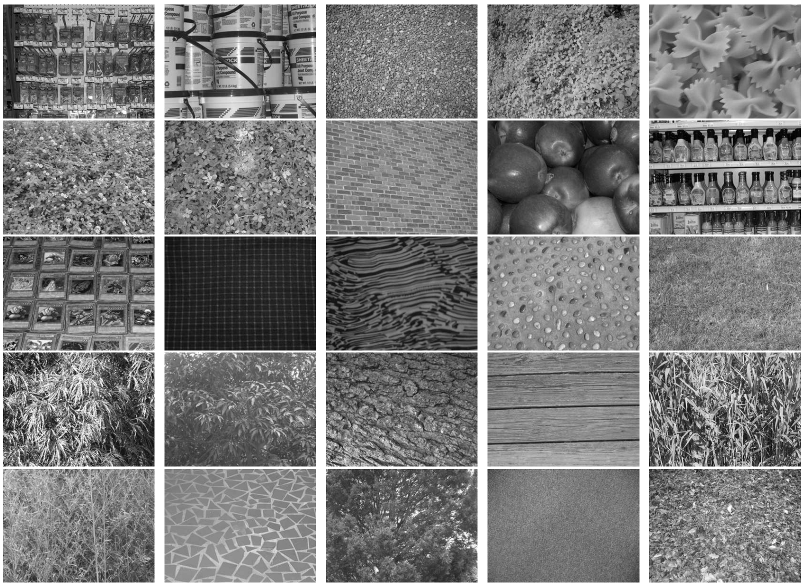


Figura 16 – Uma imagem de cada classe da base UMD.

7.1.2 Texturas UIUC

Esta base de dados foi originalmente apresentada em (LAZEBNIK et al., 2005) e apresenta um conjunto de 1000 imagens divididas em 25 classes com 40 amostras cada.

As imagens em cada classe apresentam variação de perspectiva e escala, além de terem sido coletadas sem controle na iluminação. Há diferentes tipos de objetos, como casca de madeira, tijolos, tecidos, entre outros. A Figura 17 ilustra uma imagem de cada classe. O tamanho de cada imagem original é 640×480 em tons de cinza, mas aqui também foi reduzido para 320×240.

7.1.3 Texturas KTHTIPS-2b

O banco de imagens KTHTIPS-2b (HAYMAN et al., 2004) possui um conjunto de 11 classes. Cada classe é dividida em quatro partes chamadas de (*samples*), contendo 108 imagens cada, resultando no total de 4752 imagens para esta base. Cada *sample* contém imagens coletadas em configurações específicas de escala, iluminação e perspectiva.

Entre os tipos de objetos nas imagens estão: pão integral, folha de alface, madeira, entre outros. Uma imagem de cada classe é apresentada na Figura 18. Cada

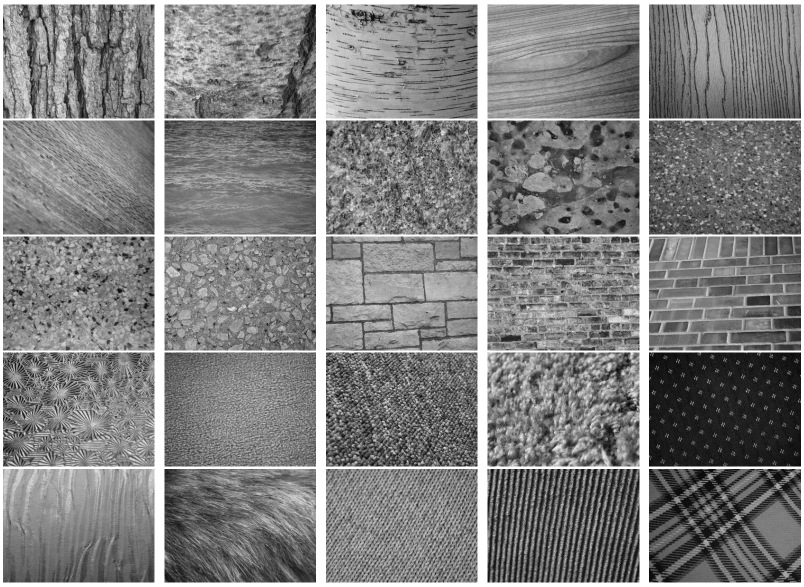


Figura 17 – Uma imagem de cada classe da base UIUC.

Figura 18 – Uma imagem de cada classe da base KTHTIPS-2b.

imagem tem o tamanho de 200 × 200. Neste caso, utilizou-se a dimensão original para a extração de descritores da base. Além disso, para que pudessem ser analisadas pelas metodologias propostas neste trabalho, as imagens foram convertidas para tons de cinza[1].

7.2 Resultados e Experimentos

Na sequência são apresentadas as taxas de acerto da classificação de imagens nas bases UMD, UIUC e KTHTIPS-2b.

[1] Foi utilizado o comando no python .convert(L), trabalhando com a luminância da imagem.

7.2.1 Resultados Textura UMD

Nesta base de dados, metade das imagens foi aleatoriamente separada para treinamento e a outra metade para testes. O resultado da taxa de acerto é a média aritmética de 10 vezes em que esta divisão aleatória de treino/teste é repetida.

Para o ELM com regressão *ridge*, temos o papel crucial do parâmetro $\gamma > 0$. Neste cenário, a taxa de classificação foi verificada para 9 neurônios na camada oculta e com diferentes valores de γ. Os resultados estão na Figura 19.

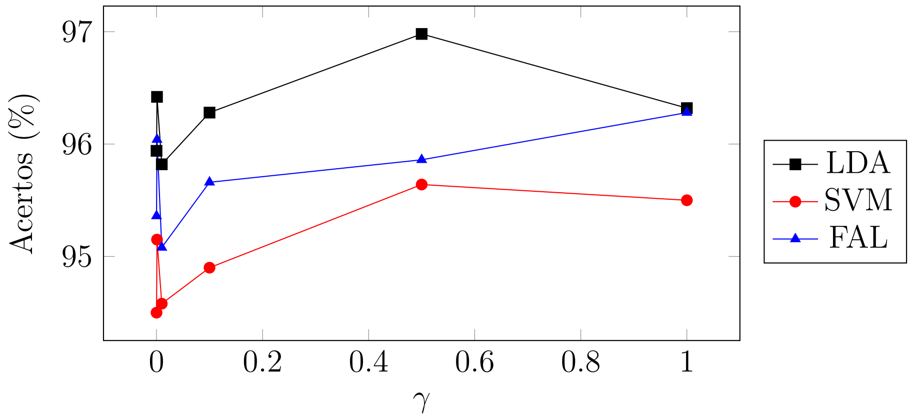


Figura 19 – Variação de γ $(10^{-4}, 10^{-3}, 10^{-2}, 10^{-1}, 0.5, 1)$ na regressão *ridge* na base UMD.

Considerando-se os valores mostrados na Figura 19, foi adotado um valor de $\gamma = 0.5$ como ótimo para a extração dos descritores. Aqui temos um fenômeno interessante. Note que pelos valores de γ encontrados na Figura 14, o erro teórico de aproximação vai para zero quando $\gamma \to 0$, porém, a melhor classificação na Figura 19 foi obtida com um valor maior de γ testado. Isso pode ser explicado pelo fato de que aqui estamos interessados na maior acurácia da classificação do conjunto de teste, que não necessariamente corresponde ao menor erro do modelo matemático aplicado para encontrar o vetor de descritores. Na realidade, o uso de um vetor que não está associado a um erro 0 na aproximação da rede ELM evita o chamado *overfitting* do modelo, quando o algoritmo se especializa no treino, mas não generaliza como o esperado nas amostras de teste.

Um parâmetro importante nos métodos analisados é o número de neurônios na camada oculta. Com isto, foram testados os seguintes valores para este parâmetro: $9, 19, 29, 39, 49$ e 59. No caso de ELM com regressão *ridge*, fixou-se o valor $\gamma = 0.5$, como definido a partir da Figura 19. Na Figura 20 são apresentados os resultados da classificação utilizando os algoritmos ELM, ELM_{LBP}, $ELM + LBP$ e ELM_{ridge}, LBP_{ELM}, $LBP_{ELM} + LBP$, ELM_{MM}, ELM_{morf}.

A Figura 20 mostra que os classificadores LDA e FAL obtiveram as melhores taxas de acerto na classificação, com exceção apenas da abordagem LBP_{ELM}. Um ponto interessante é que estes são classificadores com bom nível de interpretabilidade, já que o primeiro se baseia diretamente no Teorema de Bayes, que é inerente a modelos de classificação supervisionados, e o segundo é uma evolução da ideia de árvores de decisão, que levam em conta a influência direta de cada atributo sobre as classes no treinamento. Pode-se aferir a partir daqui então que os descritores desenvolvidos nas diversas variantes de ELM aqui propostas são de fato discriminantes efetivos por si só, sem a necessidade de complexas transformações e pré-processamentos altamente não lineares para que a classificação seja bem sucedida.

Outra constatação relevante é que para todos os métodos analisados, a partir de 39 neurônios na camada oculta, houve estabilização na taxa de acerto, fato este que do ponto de vista teórico pode ser considerado como uma consequência do Teorema 3, isto é, o nível de aproximação da rede ELM se mantém estável após um certo limiar finito (e frequentemente pequeno) do número de neurônios na camada oculta.

Observando-se os resultados apresentados na Figura 20 para a metodologia ELM_{LBP}, nota-se que a estratégia de aplicar o algoritmo ELM em imagens após a transformada LBP não apresentou melhora significativa nesta base de dados. Ou seja, a mudança no espaço do dado de entrada e saída do algoritmo ELM não contribuiu para enriquecer o vetor de descritores. Uma possível explicação para este resultado é o fato de já estarmos trabalhando com um conjunto de texturas cujas imagens originais são facilmente discriminadas, de modo que a introdução de uma transformada como LBP, que altera substancialmente a textura original, tende a ser mais prejudicial do que benéfica.

Por outro lado, os valores apresentados na Figura 20 mostram a competitividade das abordagens $ELM + LBP$ e ELM_{ridge} em relação ao algoritmo original de ELM em texturas. Ou seja, o uso de códigos LBP apenas na saída Y da rede ELM (metodologia $ELM + LBP$) tende a aumentar a performance dos descritores, já que

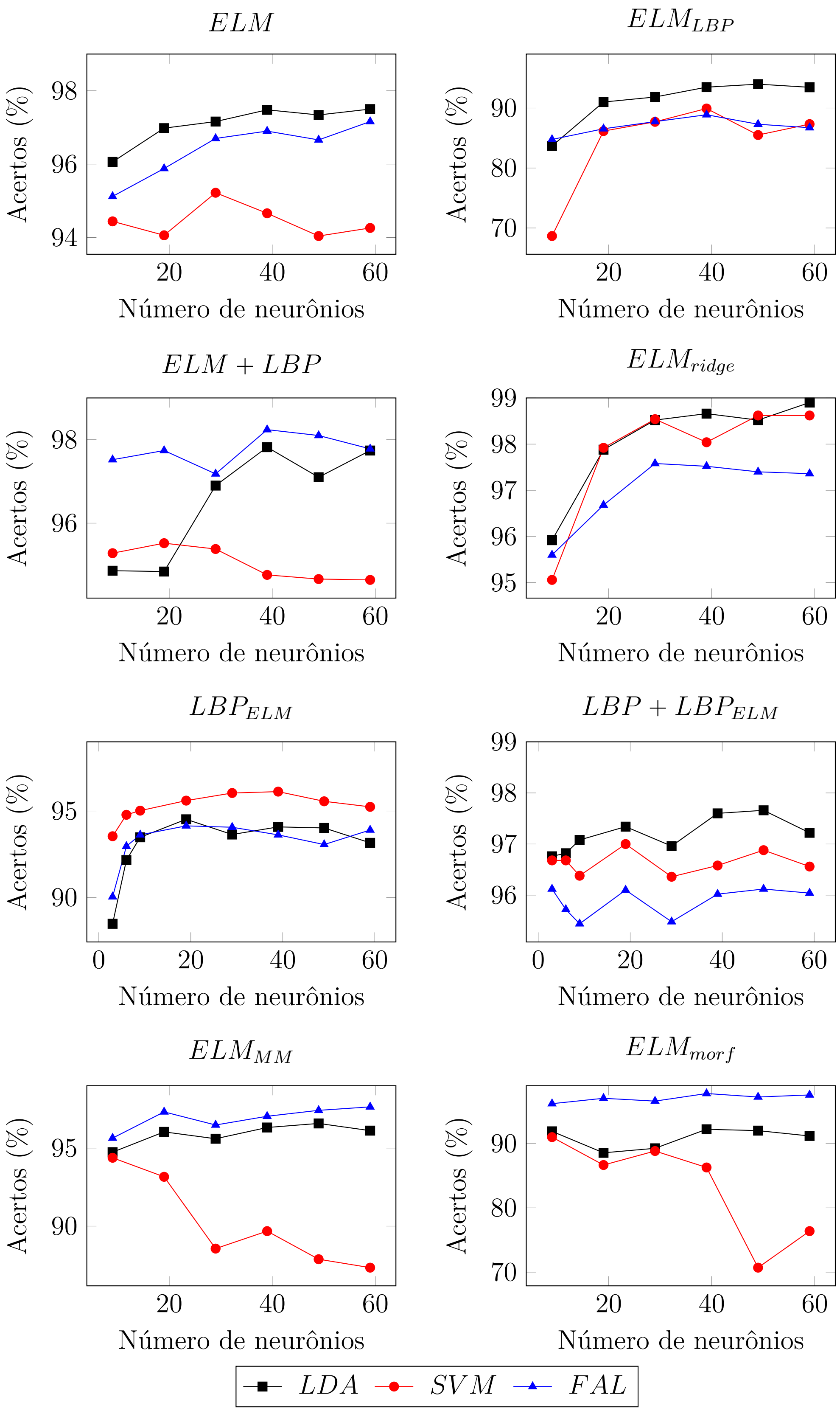


Figura 20 – Classificação da base UMD para os diferentes descritores de textura ELM analisados e variando-se o número de neurônios na camada oculta.

agora informações a respeito da influência direta dos pixels vizinhos na matriz X são também introduzidas no vetor Y.

Note-se que, diferentemente do que ocorria no método ELM_{LBP} anterior, ambas as abordagens agora possuem caráter regularizador, já que o uso de um pixel central que carrega informação dos vizinhos no método $ELM + LBP$ faz com que a função que mapeia a entrada para saída na rede ELM seja mais suave, de modo similar ao que ocorre naturalmente no algoritmo ELM_{ridge}. Vale observar ainda que temos uma melhora na acurácia possibilitada por este método mesmo que o erro de reconstrução da imagem original, ocasionado pelo modelo matemático de cálculo da matriz M na Equação 3.4 seja maior, já que não temos mais a presença explícita do pixel central no modelo.

Outra solução proposta neste trabalho foi a aplicação de LBP em imagens transformadas pelo ELM. Em geral, não houve aumento na taxa de acerto na abordagem do LBP_{ELM} no caso da base UMD. Porém, a combinação dos descritores da transformação ELM com os descritores LBP originais propiciou melhora na acurácia da classificação. A transformação ELM apresentada neste trabalho tem a função de tornar a textura mais homogênea e menos sensível a variações locais espúrias. Ao mesmo tempo, variações locais características da imagem também são atenuadas e isso justifica uma leve queda no desempenho da solução LBP_{ELM}. Por outro lado, esses descritores com características menos localizadas, quando combinados com o LBP original, acabam sendo responsáveis por um aumento geral na taxa de acerto, na medida em que mais informação da vizinhança é trazida para os descritores da textura analisada.

Por fim, temos a abordagem ELM morfológica, tanto em sua versão puramente morfológica ELM_{morf} quanto na mista ELM_{MM}. Chama atenção em ambas a estabilidade da acurácia quando se varia o número de neurônios na camada oculta. O bom desempenho com poucos neurônios pode ser justificado pela ação do operador *max* em cada neurônio. De modo similar ao que ocorre nas estratégias de *max-pooling* nas redes neurais profundas (HASTIE et al., 2001), a convolução morfológica tem o papel de dar robustez à rede, garantindo maior tolerância a micro-variações na textura. Essa mesma insensibilidade porém compromete também a capacidade do

ELM aprender detalhes finos e assim a acurácia acaba se mantendo estável. Essa é uma solução interessante no caso em que o número limitado de neurônios é interessante em função de um dado de entrada muito grande ou da disponibilidade limitada de recursos computacionais.

A Tabela 1 apresenta os melhores resultados de cada método abordado nesta dissertação para a base de dados UMD. Uma comparação geral com resultados publicados para a mesma base na literatura será ainda apresentada ao final deste capítulo, na Tabela 7.

Tabela 1 – Maior taxa de acerto para cada descritor baseado em ELM na base UMD.

Configuração	Acertos %
$(ELM)_{39}$	97.48
$(ELM_{LBP})_{59}$	93.46
$(ELM + LBP)_{39}$	98.24
$(ELM_{ridge})_{59}$	**98.90**
$(LBP_{ELM})_{39}$	96.12
$((LBP_{ELM})_{39} + LBP)$	97.60
$(ELM_{MM})_{59}$	97.64
$(ELM_{morf})_{39}$	97.78

Na Tabela 2 são apresentados os resultados mais revelantes para os testes feitos com diversas combinações (concatenações) dos descritores apresentados. O sinal de “+” representa a junção dos descritores distintos e o número $_N$ subscrito corresponde à quantidade de neurônios na camada oculta do ELM.

Com as concatenações, foi possível chegar-se às mais altas taxas de acerto na maioria dos cenários, destacando-se a junção dos quatro métodos do ELM testados nesta base de dados. Isso mostra também o caráter complementar dos descritores desenvolvidos, de modo que cada um pode contribuir com pontos de vista diferentes sobre a mesma textura e assim resultar em um modelo geral de classificação mais robusto.

Além da taxa de acerto usada até o momento, outra maneira de avaliar a qualidade da classificação é por meio da análise da matriz de confusão. A Figura 21 apresenta essa matriz para a configuração de maior taxa de acerto na Tabela 2, ou seja, a concatenação $(ELM_{ridge})_{29} + (ELM_{ridge})_{39}$ com o classificador LDA.

Tabela 2 – Concatenação de diferentes abordagens de descritores ELM de texturas apresentadas neste trabalho, para a base UMD.

Configuração	LDA %	SVM %	FAL %
$(ELM_{ridge})_9 + (ELM_{ridge})_{59}$	96.26	98.96	98.10
$(ELM_{ridge})_{19} + (ELM_{ridge})_{59}$	99.00	98.76	97.56
$(ELM_{ridge})_{29} + (ELM_{ridge})_{39}$	**99.06**	98.94	98.66
$(ELM)_9 + (ELM)_{29} + (ELM)_{29}$	98.60	96.26	97.74
$(ELM_{ridge})_9 + (ELM_{ridge})_{19} + (ELM_{ridge})_{29}$	98.92	98.96	98.20
$(ELM_{ridge})_9 + (ELM + LBP)_9$	97.60	95.52	98.10
$(ELM_{ridge})_{19} + (ELM + LBP)_{19}$	98.42	96.22	97.94
$(ELM_{ridge})_{29} + (ELM + LBP)_{29}$	98.46	96.34	97.94
$(ELM_{MM})_9 + (ELM_{morf})_9$	98.52	96.44	98.04
$(ELM)_9 + (ELM + LBP)_9 +$ $+(ELM_{ridge})_9 + (ELM_{LBP})_9$	99.00	96.30	98.52
$(ELM)_{19} + (ELM + LBP)_{19} +$ $+(ELM_{ridge})_{19} + (ELM_{LBP})_{19}$	98.54	96.60	97.80
$(ELM)_9 + (ELM + LBP)_9 +$ $+(ELM_{MM})_9 + (ELM_{morf})_9$	98.56	92.58	98.42
$ELM - LBP_9 + (ELM + LBP)_{29} + LBP$	96.60	96.16	98.48
$(ELM + LBP)_{19} + (ELM + LBP)_{29} + LBP$	96.66	95.88	98.56

Matriz de confusão

Verdadeiros

	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20	21	21	23	24	25
1	**19**	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
2	0	**22**	0	0	0	0	0	0	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
3	0	0	**19**	0	0	0	0	0	1	0	0	0	0	0	0	0	0	1	0	0	0	0	0	0	0
4	0	0	0	**25**	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
5	0	0	0	0	**24**	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
6	0	0	0	0	0	**23**	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
7	0	0	0	0	0	0	**22**	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
8	0	0	0	0	0	0	0	**19**	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
9	0	0	0	0	0	0	0	0	**22**	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
10	0	0	0	0	0	0	0	0	0	**19**	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
11	0	0	0	0	0	0	0	0	0	0	**23**	0	0	0	0	0	0	0	0	0	0	0	0	0	0
12	0	0	0	0	0	0	0	0	0	0	0	**19**	0	0	0	0	0	0	0	0	0	0	0	0	0
13	0	0	0	0	0	0	0	0	0	0	0	0	**17**	0	0	0	0	0	0	0	0	0	0	0	0
14	0	0	0	0	0	0	0	0	0	0	0	0	0	**21**	0	0	0	0	0	0	0	0	0	0	0
15	0	0	0	0	0	0	0	0	0	0	0	0	0	0	**24**	0	0	0	0	0	0	0	0	0	0
16	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	**14**	0	2	0	0	0	0	0	0	0
17	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	**17**	0	0	0	0	0	0	0	0
18	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	**17**	0	0	0	0	0	0	0
19	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	**16**	0	0	0	0	0	0
20	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	**19**	1	0	0	0	0
21	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	**16**	0	0	0	0
22	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	**17**	0	0	0
23	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	**21**	0	0
24	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	**18**	0
25	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	**21**

Classificados

Figura 21 – Matriz de confusão para a configuração $(ELM_{ridge})_{29} + (ELM_{ridge})_{39}$, na base UMD.

A matriz de confusão expressa como cada classe é discriminada pelo algoritmo. O eixo vertical indica a classe verdadeira da imagem e o horizontal a classe fornecida pelo classificador. Assim, o valor do elemento de índices ij da matriz corresponde ao número de amostras pertencentes à classe i e atribuídas pelo classificador à classe j. Por exemplo, analisando-se a Figura 21 (considerando-se a imagem como uma matriz quadrada $B_{25\times25}$), o elemento $b_{1,1} = 19$ indica que estas 19 imagens foram classificadas corretamente na classe 1. Já o valor $b_{16,18} = 2$ indica que o algoritmo classificou duas imagens que na realidade pertenciam à classe 16 como sendo da classe 18. Desta maneira, em um cenário de classificação ideal teríamos valores diferentes de zero apenas na diagonal principal. A soma dos elementos da linha corresponde a quantas imagens da classe foram separadas para teste.

Pelo fato de termos poucos elementos diferentes de zero fora da diagonal principal, pode-se afirmar que o processo de extração de descritores apresentou boa capacidade de discriminar entre as diferentes classes da base de imagens. Isso de fato já era algo esperado dada a alta acurácia já exibida anteriormente pelos descritores para a base UMD.

7.2.2 Resultados Texturas UIUC

Na base de dados UIUC, o processo de divisão treino/teste para a classificação supervisionada foi igual ao da UMD. Novamente, foram extraídos descritores usando-se as sete formas diferentes aqui estudadas: ELM original, ELM sobre transformada LBP, ELM com LBP na saída, transformada ELM, ELM com regressão *ridge* e ELM morfológico.

Os algoritmos de extração de descritores com ELM são baseados na resolução da Equação 3.5 utilizando o conceito matemático de Matriz de Moore-Penrose.

Além deste conceito, este trabalho também estudou descritores gerados a partir da resolução da Equação 3.5, com regressão *ridge*, como discutido no Capítulo 6. Primeiramente, verificou-se a acurácia na classificação para diferentes valores de γ, com o número de neurônios na camada oculta do ELM fixo (neste caso foi utilizado $N = 9$). Os resultados estão na Figura 22.

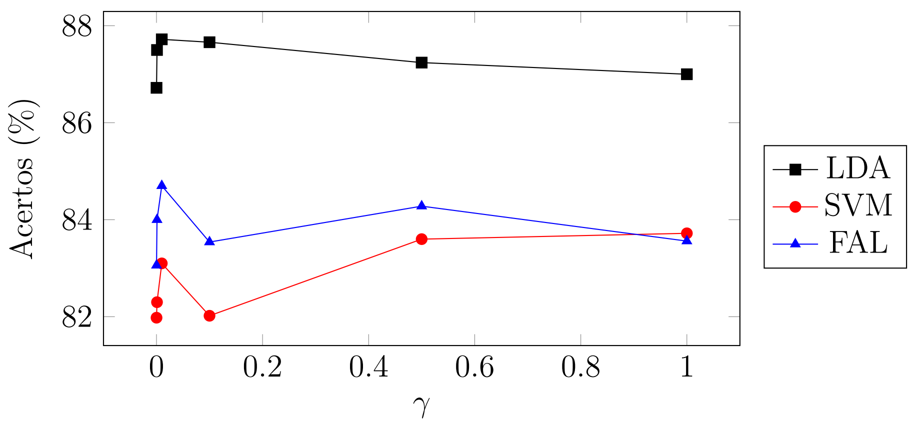


Figura 22 – Variação de γ $(10^{-4}, 10^{-3}, 10^{-2}, 10^{-1}, 0.5, 1)$ na regressão *ridge* na base UIUC.

Com base nos valores da taxa de acerto mostrados na Figura 22, optou-se aqui pela adoção do valor $\gamma = 10^{-2}$. Fixado este parâmetro, fez-se um estudo então da influência da variação do número de neurônios na camada oculta nesta abordagem. Os resultados estão na Figura 23. Na mesma Figura 23, são apresentados também os resultados da classificação para as outras cinco variantes do algoritmo ELM com diferentes números de neurônios na camada oculta.

Uma inspeção visual da Figura 23, especialmente no caso em que o classificador usado foi o LDA, mostra uma estabilização na taxa de acerto a partir de 39 neurônios na camada oculta. Este é um comportamento similar ao já observado na base UMD e a explicação para isso novamente está relacionada ao limite da capacidade de representação da rede neural do modelo ELM. Outro aspecto observado na Figura 23, e que é similar ao que ocorreu na base UMD, é que na base UIUC novamente o classificador LDA apresentou as melhores taxas de acerto entre os três classificadores avaliados.

A primeira solução aqui apresentada aplica o algoritmo de ELM sobre a transformada LBP, usando como alvo na rede o pixel original (ELM_{LBP}) e seu código LBP ($ELM + LBP$). Novamente, a segunda opção se mostra substancialmente mais precisa do que a primeira. O caráter agregador do código LBP na saída novamente se mostra relevante para os descritores propostos.

Já o método ELM_{ridge} faz uso da regularização de Tikhonov. Os resultados são bastante competitivos considerando-se que a alteração no algoritmo geral do ELM é mínima. Nota-se aqui uma característica semelhante à da base UMD, com texturas relativamente fáceis de se discriminar e consequentemente um nível alto de redundância que é atenuado pela regularização.

Outra metodologia utilizada baseia-se na transformação ELM proposta neste trabalho. Os descritores neste caso foram obtidos a partir de códigos LBP calculados sobre a transformada ELM. Não houve melhora com o uso apenas da transformada ELM em relação aos descritores LBP da imagem original, mas a combinação de ambos os descritores melhorou sensivelmente a taxa de acerto, quando esta é comparada à versão original do LBP, mostrando que a reconstrução da imagem a partir do vetor de pesos da rede ELM contribui para uma representação mais rica e robusta

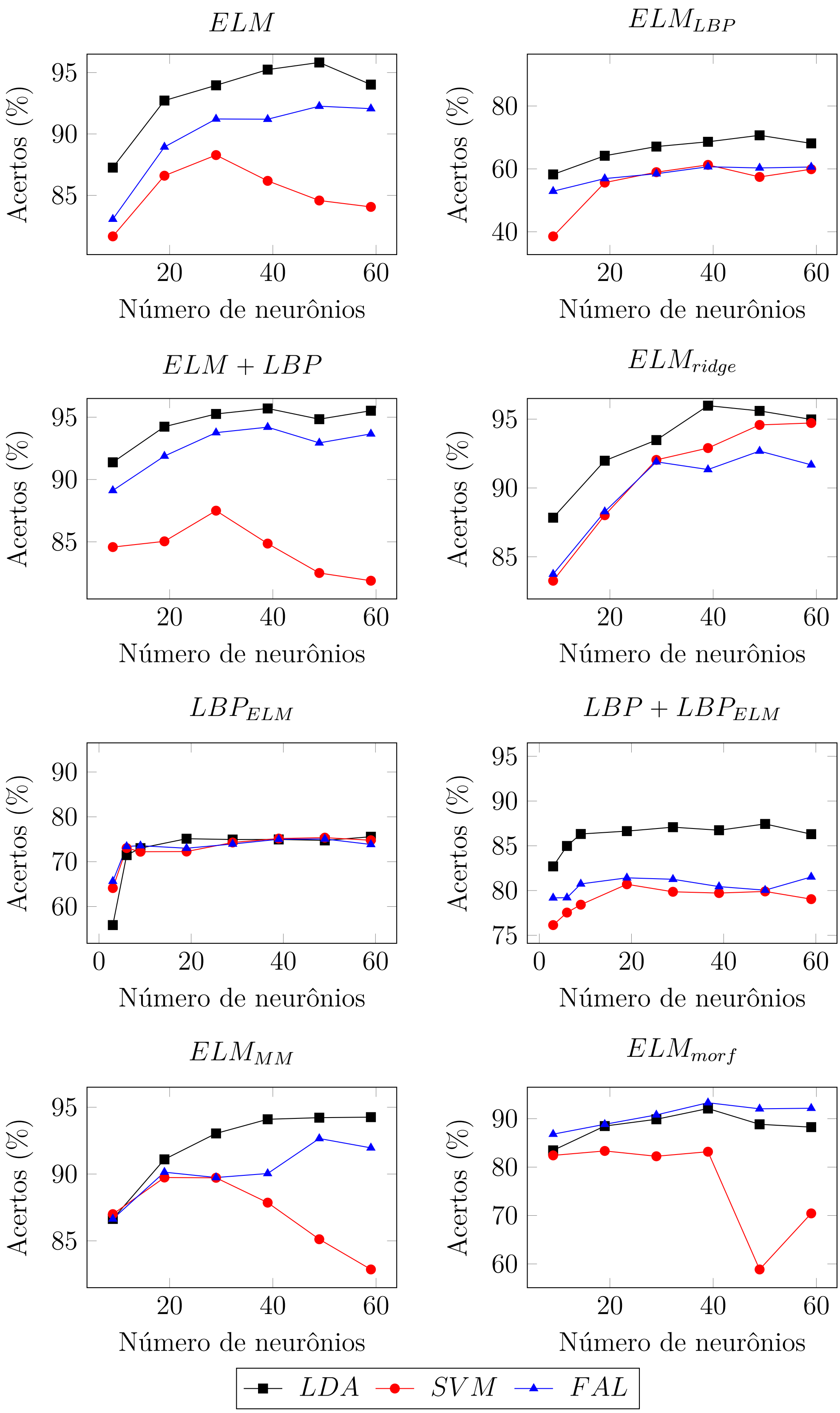


Figura 23 – Classificação da base UIUC para os diferentes descritores de textura ELM analisados e variando-se o número de neurônios na camada oculta.

da textura original.

Por fim, temos a representação morfológica da rede ELM. O método misto (ELM_{MM}) se sobressaiu, confirmando que neurônios morfológicos têm papel mais relevante na geração de descritores quando são combinados com neurônios clássicos.

Os melhores resultados de cada descritor ELM são apresentados na Tabela 3 e uma comparação com a literatura se encontra na Tabela 7 ao final deste capítulo. O método com regularização *ridge* se sobressaiu em relação às demais variantes de ELM estudadas neste projeto.

Tabela 3 – Maior taxa de acerto para cada descritor baseado em ELM na base UIUC.

Configuração	Acertos %
$(ELM)_{49}$	95.82
$(ELM_{LBP})_{49}$	70.68
$(ELM + LBP)_{39}$	95.70
$(ELM_{ridge})_{39}$	**95.98**
$(LBP_{ELM})_{59}$	75.56
$((LBP_{ELM})_{49} + LBP)$	87.44
$(ELM_{MM})_{49}$	94.22
$(ELM_{morf})_{39}$	93.30

Fazendo uma comparação entre as Tabelas 1 e 3, pois as bases possuem o mesmo número de imagens, classes e tamanho. Nesta base houve ganho menor do ELM com regressão *ridge* sobre o ELM padrão e as porcentagens de acertos também ficaram inferiores nos outros métodos. Porém, neste caso, as imagens possui variação na perspectiva e escala, aumentando a dificuldade para a extração de descritores e as taxas de acertos estão superiores a 90 % nos melhores testes, ou seja, já é notável uma boa classificação das imagens.

Além da classificação apresentada na Figura 23 e Tabela 3, também foram combinadas diferentes configurações e métodos de descritores ELM. Os principais resultados testados são apresentados na Tabela 4. De acordo com essa tabela, houve um pequeno aumento na taxa de acerto com as combinações dessas abordagens, com destaque para a concatenação de descritores usando LBP na saída do modelo ELM ($ELM + LBP$) com os obtidos por regressão *ridge*.

Seguindo o mesmo protocolo adotado para a base de dados UMD, aqui também

Tabela 4 – Concatenação de diferentes abordagens de descritores ELM de texturas apresentadas neste trabalho, para a base UIUC.

Configuração	LDA %	SVM %	FAL %
$(ELM)_9 + (ELM)_{19}$	95.34	90.32	92.84
$(ELM)_{19} + (ELM)_{29}$	95.62	91.18	93.22
$(ELM + LBP)_9 + (ELM + LBP)_{19}$	96.10	89.88	95.44
$(ELM + LBP)_{19} + (ELM + LBP)_{29}$	96.42	89.12	95.08
$(ELM_{ridge})_{19} + (ELM + LBP)_{19}$	96.00	87.02	92.80
$(ELM_{ridge})_{29} + (ELM + LBP)_{29}$	**96.84**	89.90	95.04
$(ELM_{ridge})_{39} + (ELM + LBP)_{39}$	96.42	85.84	94.60
$(ELM + LBP)_9 + (ELM + LBP)_{19}+$ $+(ELM + LBP)_{29}$	96.28	90.28	95.96

foi analisada a matriz de confusão da melhor configuração de descritor ELM. A Figura 24 apresenta a matriz para a configuração $(ELM_{ridge})_{29} + (ELM + LBP)_{29}$.

As matrizes de confusão das Figuras 21 e 24 são semelhantes, fato este que é uma consequência das características similares das bases de dados e maneiras como a divisão treino/teste é executada.

7.2.3 Resultados na Base KTHTIPS-2b

Nesta base, o modelo de divisão treino/teste foi diferente do usado na UIUC e UMD, já que agora cada classe tem uma divisão interna própria em quatro subconjuntos (*samples*) com 108 imagens cada.

Para treinamento, foi usado um *sample* de cada classe e foram utilizados os outros três *samples* para teste. Este procedimento foi repetido de modo que cada *sample* de imagens fosse usado uma vez para treinamento, resultando ao todo em quatro combinações treino/teste para cada configuração de descritores. Ao final,

Matriz de confusão

Verdadeiros

1	**20**	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
2	0	**24**	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
3	3	0	**21**	0	1	0	0	0	0	0	0	0	1	0	0	0	0	0	1	0	0	0	0	0	0
4	0	0	0	**14**	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	1	0	1	0	0	0
5	0	0	0	0	**18**	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
6	0	0	0	0	0	**20**	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
7	0	0	0	0	0	0	**20**	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
8	0	0	0	0	0	0	0	**20**	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
9	0	0	0	0	0	0	0	0	**19**	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
10	0	0	0	0	0	0	0	0	0	**16**	0	0	0	0	0	0	0	0	1	0	0	0	0	0	0
11	0	0	0	0	0	0	0	0	0	0	**22**	0	0	0	0	0	0	0	0	0	0	0	0	0	0
12	0	0	0	0	0	0	0	0	0	0	0	**17**	0	0	0	0	0	0	0	0	0	0	0	0	0
13	0	0	0	0	0	0	0	0	0	0	0	0	**18**	0	0	0	0	0	0	0	0	0	0	0	0
14	0	0	0	0	0	0	0	0	0	0	0	0	0	**17**	0	0	0	0	0	0	0	0	0	0	0
15	0	0	0	0	0	0	0	0	0	0	0	0	0	1	**20**	0	0	0	0	0	0	0	0	0	0
16	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	**17**	0	0	0	0	0	0	0	0	0
17	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	**23**	0	0	0	0	0	0	0	0
18	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	**23**	1	0	0	0	0	0	0
19	0	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	1	**21**	0	0	0	0	0	0
20	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	**20**	0	0	0	0	0
21	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	**25**	0	0	0	0
22	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	**13**	0	0	0
23	0	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	1	0	0	0	**15**	0	0
24	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	**21**	0
25	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	1	0	0	**19**
	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20	21	21	23	24	25

Classificados

Figura 24 – Matriz de confusão para a configuração $(ELM_{ridge})_{29} + (ELM + LBP)_{29}$, na base UIUC.

calculou-se a média aritmética dos acertos para cada método de classificação.

No caso do ELM com regressão *ridge*, fez-se primeiramente uma análise da influência do valor de γ para a extração dos descritores na taxa de acerto. Neste caso, foi usado um valor fixo de 9 neurônios na camada oculta e os resultados estão na Figura 25. Pelos valores encontrados na Figura 25, optou-se por adotar o valor $\gamma = 1.0$ para que se possa em seguida variar o número de neurônios na camada oculta do ELM_{ridge}.

Por sua vez, a Figura 26 apresenta os resultados da classificação para diferentes configurações de ELM estudadas, variando-se o número de neurônios na camada oculta. Devido ao alto custo computacional exigido para esta base, consequência

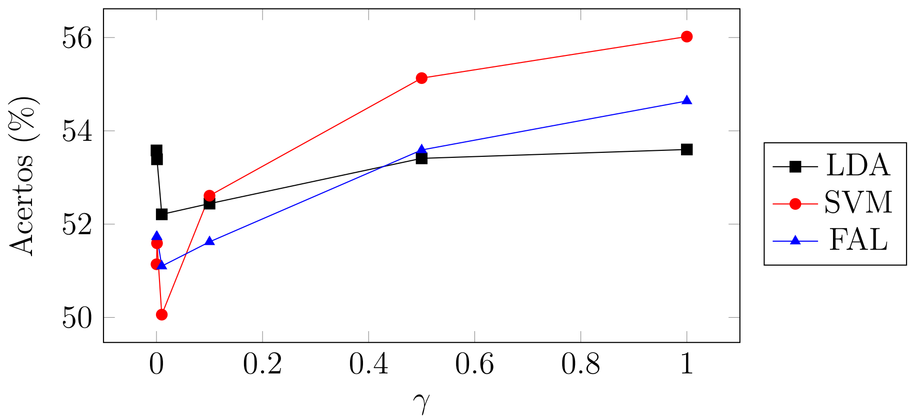


Figura 25 – Variação de γ $(10^{-4}, 10^{-3}, 10^{-2}, 10^{-1}, 0.5, 1,)$ na regressão *ridge* na base KTHTIPS-2b.

do grande número de imagens, foram testadas configurações com até 59 neurônios na camada oculta, percebendo-se que houve uma estabilização na porcentagem de acerto a partir disso.

Uma análise da Figura 26 mostra que, de modo análogo ao que ocorreu nos testes nas bases UMD e UIUC, houve aqui também uma estabilização na taxa de acerto. Neste caso, em boa parte dos testes, a partir de 19 neurônios na camada oculta já não se observou mais uma melhora significativa da acurácia na classificação. Essa estabilização precoce está associada ao fato de as imagens terem tamanho menor do que nas bases anteriores, o que acarreta em menor quantidade de informação para o modelo do ELM aprender durante o processo e maior tendência a *overfitting* quando um número alto de neurônios é empregado na camada oculta. Nota-se nesta base também um melhor desempenho do classificador florestas aleatórias, algo também condizente com a alta complexidade e não linearidade dos padrões de textura característicos dessas imagens.

De um modo geral e semelhantemente ao que se observou nas demais bases, as abordagens de ELM com regressão *ridge*, $ELM + LBP$ e ELM_{MM} alcançaram os melhores resultados e se mostraram bastante competitivas com o algoritmo de ELM em texturas original.

Outra abordagem explorada foi a transformação ELM e o uso de descritores LBP a partir dessa transformada. A concatenação dos descritores de LBP sobre a imagem original e sobre a transformada ELM seguiu o padrão dos resultados nas

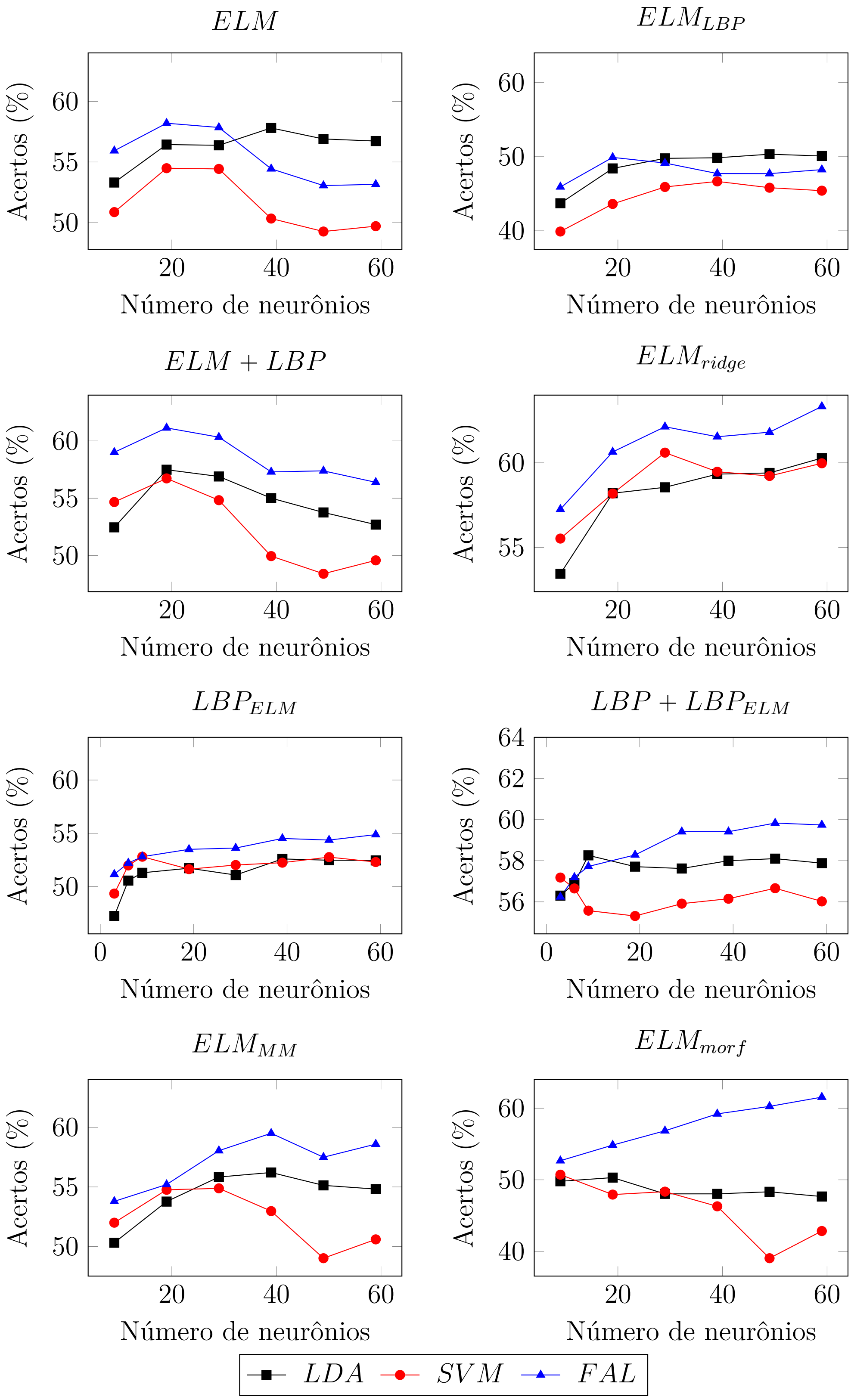


Figura 26 – Classificação da base KTHTIPS-2b para os diferentes descritores de textura ELM analisados e variando-se o número de neurônios na camada oculta.

bases UMD e UIUC, de modo que essa combinação de descritores sobre dois espaços de representação diferentes aumentou a porcentagem de acerto também nesta base.

A Tabela 5 destaca os melhores resultados de cada método, sendo que uma comparação com trabalhos que utilizaram a mesma metodologia de treino e teste na literatura é apresentada na Tabela 7.

Tabela 5 – Maior taxa de acerto para cada descritor baseado em ELM na base KTHTIPS-2b.

Configuração	Acertos %
$(ELM)_{19}$	58.20
$(ELM_{LBP})_{49}$	50.33
$(ELM + LBP)_{19}$	61.14
$(ELM_{ridge})_{59}$	**63.33**
$(LBP_{ELM})_{39}$	54.52
$((LBP_{ELM})_{49} + LBP)$	59.83
$(ELM_{MM})_{39}$	59.50
$(ELM_{morf})_{59}$	61.55

Além dos testes anteriores, foram exploradas ainda diferentes combinações dos métodos propostos, como apresentado na Tabela 6. De acordo com a última linha da Tabela 6, uma combinação de quatro abordagens aqui trabalhadas para o ELM (ELM original, $ELM + LBP$, ELM_{ridge} e ELM_{LBP}) apresentou os melhores resultados em relação à taxa de acerto nesta base.

Para a melhor configuração na Tabela 6, a matriz de confusão correspondente é apresentada na Figura 27. Observando-se a Figura 27 e a Tabela 6, constata-se que essa base de imagens apresentou os resultados mais baixos na porcentagem de acerto da classificação, quando comparada com UIUC e UMD. Vale destacar entretanto que aqui foram utilizados para treino apenas 25% das imagens, enquanto UMD e UIUC usaram 50%. Além disso, a subdivisão em *samples* torna o problema da classificação desta base sensivelmente mais desafiador, já que cada *sample* corresponde a configurações substancialmente distintas de iluminação, escala e perspectiva. O fato de que o treinamento é feito usando apenas um desses *samples* exige que o algoritmo seja capaz de aprender padrões para o teste que não viu na fase de treino. Por fim, as abordagens estudadas neste projeto são todas adaptadas para análise de imagens em tons de cinza, enquanto a KTHTIPS-2b é uma base de texturas

Tabela 6 – Concatenação de diferentes abordagens de descritores ELM de texturas apresentadas neste trabalho, para a base KTHTIPS-2b.

Configuração	LDA %	SVM %	FAL %
$(ELM_{ridge})_{19} + (ELM_{LBP})_{19}$	61.19	57.25	62.43
$(ELM_{ridge})_{29} + (ELM_{LBP})_{29}$	62.01	54.77	60.84
$(ELM_{ridge})_{39} + (ELM_{LBP})_{39}$	63.71	50.01	57.98
$(ELM_{ridge})_{19} + (ELM_{ridge})_{59}$	62.42	60.47	61.12
$(ELM_{ridge})_{29} + (ELM_{ridge})_{59}$	62.93	61.60	63.40
$(ELM)_9 + (ELM + LBP)_9 + $ $+ (ELM_{ridge})_9 + (ELM_{LBP})_9$	62.00	58.43	62.72
$(ELM)_{19} + (ELM + LBP)_{19} + $ $+ (ELM_{ridge})_{19} + (ELM_{LBP})_{19}$	**65.48**	58.42	63.40
$(ELM)_{29} + (ELM + LBP)_{29} + $ $+ (ELM_{ridge})_{29} + (ELM_{LBP})_{29}$	65.10	55.34	59.77

coloridas. Os métodos que atingem os melhores resultados na literatura fazem uso dessa informação trazida pelos padrões de cores nas imagens.

7.3 Observações Finais sobre as Bases de *Benchmark*

A Tabela 7 apresenta resultados disponíveis na literatura em trabalhos que utilizaram os mesmos protocolos de treino/teste bem como as mesmas versões das bases UMD, UIUC e KTHTIPS-2b adotadas neste trabalho. Também são acrescentados os melhores resultados encontrados para as abordagens propostas e/ou exploradas neste estudo.

Os dados apresentados na Tabela 7 permitem uma rápida comparação direta entre os métodos explorados neste trabalho e técnicas publicadas na literatura. Note-se por exemplo o caso de algumas dessas soluções da literatura a título de exemplo:

Tabela 7 – Taxas de acerto (%) de métodos clássicos e do estado-da-arte para classificação de texturas nas bases de dados de *benchmark*, em comparação com as abordagens aqui estudadas.

Método	UMD	UIUC	KTHTIPS-2b
MFS [1]	93.9	97.2	-
VZ-MR8 [2]	-	92.9	46.3
LBP [3]	96.1	88.4	50.3
VZ-Joint [4]	-	78.4	53.3
BSIF [5]	96.1	73.3	54.3
CLBP [6]	98.6	95.7	57.3
L. Liu et. al – CS [7]	99.1	96.3	-
PLS [8]	99.0	96.6	-
SIFT+LLC [9]	98.4	96.3	57.6
SIFT + KCB [10]	98.0	91.4	58.3
SIFT + BoVW [11]	98.1	96.1	58.4
LBP_{riu2}/VAR [12]	95.9	84.4	58.5
PCANet (NNC) [13]	90.5	57.7	59.4
RandNet (NNC) [14]	90.9	56.6	60.7
SIFT + VLAD [15]	**99.3**	96.5	63.1
ScatNet (NNC) [16]	93.4	88.6	63.7
Xu et al.-OTF [17]	98.5	**97.4**	-
DeCAF [18]	96.4	94.2	70.7
SIFT + FV [19]	-	-	
FC-CNN VGGM [20]	97.2	94.5	70.8
FC-CNN AlexNet [21]	95.9	91.1	**71.5**
(H+L)(S+R) [22]	96.9	97.0	-
(ELM)	97.5	95.8	58.2
(ELM_{LBP})	93.5	70.7	50.3
$(ELM + LBP)$	**98.2**	**95.7**	**61.1**
(ELM_{ridge})	**98.9**	**96.0**	**63.3**
$(LBP_{ELM} + LBP)$	97.6	87.4	59.8
(ELM_{MM})	97.6	94.2	59.5
(ELM_{morf})	97.8	93.3	61.5
Combinações (ELM)	**99.1**	**96.8**	**65.4**

[1](XU et al., 2009b), [2] (VARMA; ZISSERMAN, 2005), [3] (OJALA et al., 2002), [4] (VARMA; ZISSERMAN, 2009), [5] (KANNALA; RAHTU, 2012), [6] (GUO et al., 2010), [7] (LIU et al., 2011b), [8] (QUAN et al., 2014), [9] (CIMPOI et al., 2014), [10] (CIMPOI et al., 2014), [11] (CIMPOI et al., 2014), [12] (OJALA et al., 2002), [13] (CHAN et al., 2015), [14] (CHAN et al., 2015), [15] (CIMPOI et al., 2014), [16] (BRUNA; MALLAT, 2013), [17] (XU et al., 2009a), [18] (CIMPOI et al., 2014), [19] (CIMPOI et al., 2016), [20] (CIMPOI et al., 2016), [21] (CIMPOI et al., 2016), [22] (LAZEBNIK et al., 2005).

Matriz de confusão

Verdadeiros \ Classificados	1	2	3	4	5	6	7	8	9	10	11
1	**321**	0	0	1	0	1	0	0	0	0	1
2	0	**286**	0	0	8	30	0	0	1	0	0
3	0	1	**6**	74	46	87	67	28	6	8	1
4	0	0	0	**306**	0	15	2	0	1	0	0
5	0	16	15	81	**76**	14	14	41	17	45	5
6	0	0	7	57	0	**240**	12	0	1	0	7
7	2	2	0	0	0	43	**265**	0	7	5	0
8	1	15	4	8	22	3	17	**214**	0	4	36
9	0	13	0	0	0	20	5	0	**284**	2	0
10	0	0	0	1	0	2	19	0	11	**291**	0
11	0	8	67	1	30	122	8	61	5	0	**22**

Figura 27 – Matriz de confusão para a configuração $(ELM)_{19} + (ELM + LBP)_{19} + (ELM_{ridge})_{19} + (ELM_{LBP})_{19}$ na base KTHTIPS-2b.

- Abordagens ELM *versus* SIFT + KCB [10]: melhor desempenho de $(ELM + LBP)$, (ELM_{ridge}) e ELM combinado na base UIUC, melhor desempenho de todas as abordagens ELM, com exceção apenas do (ELM_{LBP}) e $LBP_{ELM}+LBP$ na UMD e melhor desempenho de $(ELM + LBP)$, (ELM_{ridge}), $(LBP_{ELM} + LBP)$ e Combinações (ELM) na KTHTIPS-2b;

- Abordagens ELM *versus* LBP [3]: melhor desempenho de todas as abordagens ELM, com exceção apenas do (ELM_{LBP}) na UMD, melhor desempenho de todas as abordagens ELM, com exceção apenas do (ELM_{LBP}) e $(LBP_{ELM} + LBP)$ na UIUC e melhor resultado de todas as abordagens ELM na KTHTIPS-2b.

- Abordagens ELM *versus* RandNet (NNC) [14]: a abordagem explorada neste trabalho alcançou melhores resultados nas bases UMD e UIUC e os métodos $(ELM+LBP)$, (ELM_{ridge}) e as combinações se sobressaíram na KTHTIPS-2b;

- Abordagens ELM *versus* SIFT + VLAD [15]: ELM teve desempenho pior na UMD, mas obteve porcentagens de acerto competitivas tanto na UIUC quanto na KTHTIPS-2b, sobretudo para as combinações de descritores ELM;

- Abordagens *ELM versus* BSIF [5]: melhores resultados da metodologia proposta em todas as bases, com exceção apenas do (ELM_{LBP}) na UIUC e KTHTIPS-2b;

- Abordagens *ELM versus* FC-CNN VGGM [20] e FC-CNN AlexNet [21]: essa é uma comparação bastante interessante pois essas abordagens da literatura são técnicas que usam redes neurais convolucionais, *transfer learning* e todo o paradigma amplamente difundido de *deep learning*. São técnicas caras computacionalmente e de difícil interpretabilidade, mas que sabidamente costumam alcançar os melhores resultados nos *benchmarks* mais populares. Aqui porém nota-se que, embora o desempenho das CNNs na KTHTIPS-2b seja efetivamente melhor (mesmo que nosso método ainda se mostre competitivo), no caso da UIUC e UMD, a metodologia baseada em *ELM* alcançou resultados melhores, mesmo usando um modelo sensivelmente mais simples em termos computacionais.

Por fim, cabe destacar que a comparação desenvolvida nesta seção está focada na taxa de acerto de cada método, o que claramente é um fator preponderante na escolha de um algoritmo de aprendizado e modelagem para um problema específico. Existem no entanto outros parâmetros, como por exemplo o tempo computacional, que também podem ser relevantes nesta decisão. Neste sentido, técnicas como as baseadas em CNN ou SIFT na Tabela 7 também se mostram desvantajosas. Em suma, os resultados encontrados trazem a percepção de que as metodologias aqui propostas apresentam resultados altamente competitivos quanto comparadas ao que existe no estado-da-arte da literatura, comprovando assim a sua eficiência como ferramenta para classificação de imagens de texturas.

8 Aplicação em Imagens Médicas

Este capítulo elucida as aplicações dos métodos de análise de imagens, descritos no capítulo 5, a problemas na área de imagens médicas.

8.1 Câncer de Pulmão

O câncer de pulmão é um dos tipos de câncer mais letais e mais frequentes em todo o mundo. É o câncer que mais mata entre homens e o segundo entre mulheres (atrás apenas do câncer de mama) (WANG et al., 2018). Este câncer pode ainda ser subdividido em dois grandes grupos: de pequenas células e de não pequenas células[1]. Entre estes dois tipos, o de pequenas células é significativamente mais agressivo.

O câncer de pulmão neuroendócrino de pequenas células é um exemplo de neoplasia que é muito agressivo desde os estágios iniciais. Nestes tumores, as células neoplásicas invadem desde o início o tecido circundante pelo movimento celular ativo, destroem o tecido original e têm acesso aos vasos linfáticos e sanguíneos. De lá, eles são levados para órgãos distantes e começam a destruir vários órgãos do hospedeiro em múltiplos focos até que o paciente morra.

A sobrevida mediana de pacientes que sofrem com este tipo de câncer em nosso país é de aproximadamente seis a oito meses. Mas é interessante notar que alguns pacientes têm uma sobrevida muito curta (menos de um mês) enquanto outros resistem por até um ano, mas muito poucos sobrevivem além disso. Cada carcinoma de pequenas células é, no momento do diagnóstico, já composto por diferentes interações de sub-clones com características fisiológicas distintas e diferentes capacidades de invasão e metastização. Essa interação de diferentes clones não é linear e pode-se presumir que modelos avançados de aprendizado de máquinas sejam adequados para descrever o comportamento deste tipo de tumor.

Distinguir entre o câncer de pequenas células e de não pequenas células é também fundamental do ponto de vista do tratamento a ser aplicado e do prognóstico

[1] Existem variantes que não se encaixam bem em nenhum destes grupos, mas são muito raras.

como um todo. O tipo de não pequenas células pode ser tratado com cirurgia, enquanto o de pequenas células requer quimio/radioterapia. Porém, mesmo no caso do não pequenas células, a abordagem cirúrgica só é viável se identificado nos estágios iniciais.

Tudo isso reforça o interesse em um diagnóstico preciso e o mais precoce possível. E as ferramentas do aprendizado de máquinas têm sido cada vez mais investigadas como um auxílio ao patologista no diagnóstico deste tipo de doença. Entre as vantagens da solução computacional, pode-se citar a maior rapidez, possibilidade de se treinar o algoritmo com uma junta composta pelos maiores especialistas do mundo, ausência de subjetividade, custo do processo como um todo reduzido, etc.

Dessa forma, a metodologia proposta foi aplicada ao problema da identificação da presença ou não de câncer de pulmão e, se de fato houver um câncer, qual o tipo deste câncer. Quatro grupos de amostras foram usados para validação do método: um grupo de controle (sem a doença), um grupo de pequenas células, um grupo de não pequenas células sub-tipo adenocarcinoma e um grupo de não pequenas células sub-tipo epidermoide.

As imagens usadas foram obtidas a partir de exames citológicos (células avulsas). Essa abordagem traz algumas vantagens, como o fato de ter um processo de coleta menos agressivo. Porém, a dificuldade na identificação do câncer olhando apenas para células avulsas é muito maior do que observando o tecido como um todo. Neste cenário, o uso da inteligência computacional se mostra ainda mais imprescindível para se alcançar qualquer resultado minimamente aceitável para fins práticos.

8.2 Imagens de Textura de Câncer de Pulmão

A base de imagens utilizada neste trabalho foi disponibilizada pela Faculdade de Ciências Médicas (FCM) da Unicamp. O número total de imagens é de 5277, divididas em quatro classes, sendo uma de controle sem a doença (692 imagens), outro conjunto referente ao câncer do tipo pequenas células (3572 imagens) e um grupo de imagens do tipo de câncer de não pequenas células, dividida em duas sub-classes: adenocarcinoma (529 imagens) e epidermoide (484 imagens). O estudo foi aprovado pelo comitê de ética local, sob registros CAAE: 0515.0.146.000-06 e 89409118.7.0000.5404.

As imagens são obtidas a partir de exames citológicos de 132 pacientes. O equipamento usado na coleta foi um microscópio Kontron Zeiss KS-300 e as lâminas foram digitalizadas no formato de *bit map* (.bmp), como descrito em (ADAM et al., 2006). O processo resultou ao final em um conjunto de imagens coloridas (corante Hematoxilina-Eosina), com tamanho de 256×256 e níveis de luminância variando entre 0 e 255. Estas imagens foram classificadas de maneira independente por dois patologistas, sendo que foram utilizadas para a composição da base apenas as amostras em que ambos tiveram a mesma opinião a respeito do diagnóstico.

Para a extração dos vetores de características aqui estudados, foi desconsiderado o fundo branco em volta da célula e utilizou-se apenas o retângulo de maior área tal que todos os pixels estejam dentro da região de interesse no núcleo da célula, evitando assim que informações espúrias sejam introduzidas no modelo. Nesta base, as imagens também foram convertidas para tons de cinza. A Figura 28 mostras três imagens de exemplo para cada classe (cada coluna da figura representa uma classe).

8.3 Resultados na Base de Imagens Médicas

Como vimos, esta base de dados contém quatro classes e temos também quatro formas diferentes para a análise do problema que são de interesse prático do ponto de vista médico:

- Duas classes: câncer *versus* sem câncer;
- Duas classes: pequenas células *versus* não pequenas células;
- Duas classes: adenocarcinoma *versus* epidermoide;
- Quatro classes: adenocarcinoma *versus* epidermoide *versus* pequenas células *versus* sem câncer.

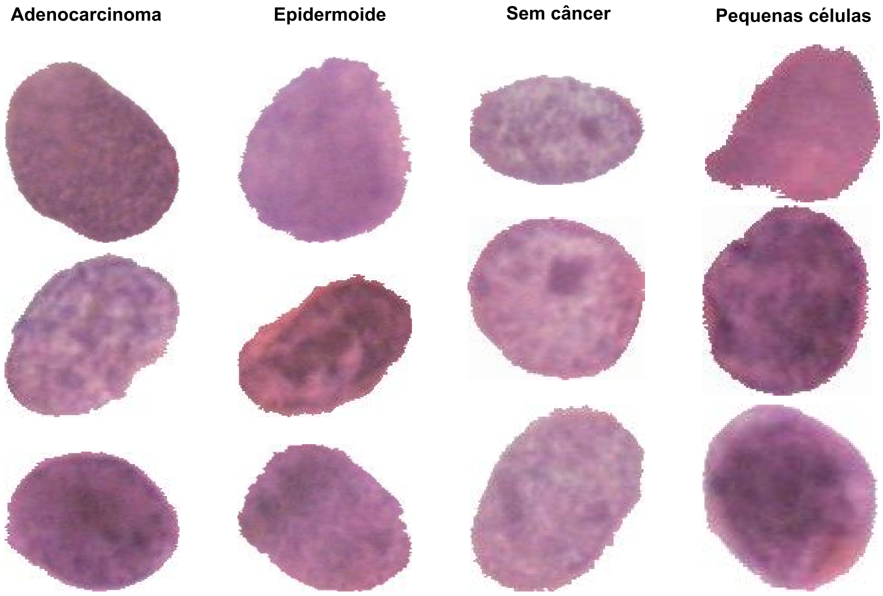


Figura 28 – Exemplos de imagens de células do pulmão, cada coluna representa uma classe.

Para a utilização efetiva dos classificadores, foram separadas aleatoriamente 90% das imagens para treinamento e 10% para teste. As medidas de desempenho apresentadas em cada uma das tabelas e gráficos a seguir representam uma média aritmética de 10 vezes em que esta simulação aleatória é repetida com variações na divisão treino/teste.

8.3.1 Normal *versus* Câncer

A Figura 30 apresenta os resultados referentes à extração de descritores e classificação para o problema de duas classes com *versus* sem câncer. O valor de $\gamma = 0.5$ para o método de regressão *ridge* foi adotado com base na Figura 29. Optou-se neste e nos resultados seguintes em usar-se "área sob a curva de *precision/recall*" em vez da taxa de acerto como antes. Isto se deve ao fato de aqui as classes serem desbalanceadas, o que geraria uma distorção no uso direto da acurácia. Como se sabe da teoria estatística, medidas como *precision/recall* (PR) e curva ROC são mais adequadas a este cenário. Na prática, essa medida de área pode ser interpretada de modo similar à ideia de acurácia. Quanto mais próximo de 1, melhor é o desempenho e mais confiável é o classificador.

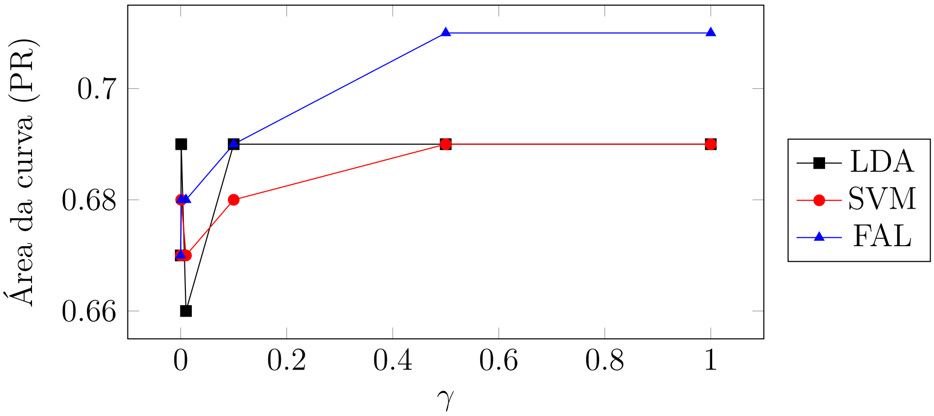


Figura 29 – Variação de γ $(10^{-4}, 10^{-3}, 10^{-2}, 10^{-1}, 0.5, 1)$ para o problema normal *versus* câncer.

Nota-se aqui que, em geral, um aumento no número de neurônios não contribuiu significativamente para a melhora do desempenho do classificador. Pode ser observado que, mesmo para 9 neurônios na camada oculta, já são alcançados valores próximos da área máxima. A relativa facilidade da classificação neste problema é um dos principais fatores para isso pois o aumento de neurônios tende a manter ruídos e detalhes espúrios das imagens originais, que degradam a capacidade do classificador.

Percebe-se ainda que novamente o método ELM_{ridge} se sobressaiu frente às demais variações, como já havia sido observado nas bases de *benchmark*. De fato, a sutileza de detalhes que permite a discriminação entre essas duas categorias é altamente sensível a ruídos inerentes ao processo de captura dessas imagens e a introdução de um termo regularizador explícito no processo ajuda a atenuar esse efeito e consequentemente aumentar a acurácia do classificador.

Em relação às metodologias baseadas na transformação ELM, neste experimento a abordagem proposta já apresentou bons resultados mesmo sem a combinação com os descritores LBP originais, diferentemente do que foi apresentado nas bases de *benchmark*. Novamente, porém, a combinação com o LBP sobre a imagem original propiciou um desempenho ainda melhor, sobretudo quando um número reduzido de neurônios na camada oculta foi empregado.

A concatenação de diferentes configurações é apresentada na Tabela 8. Nota-se

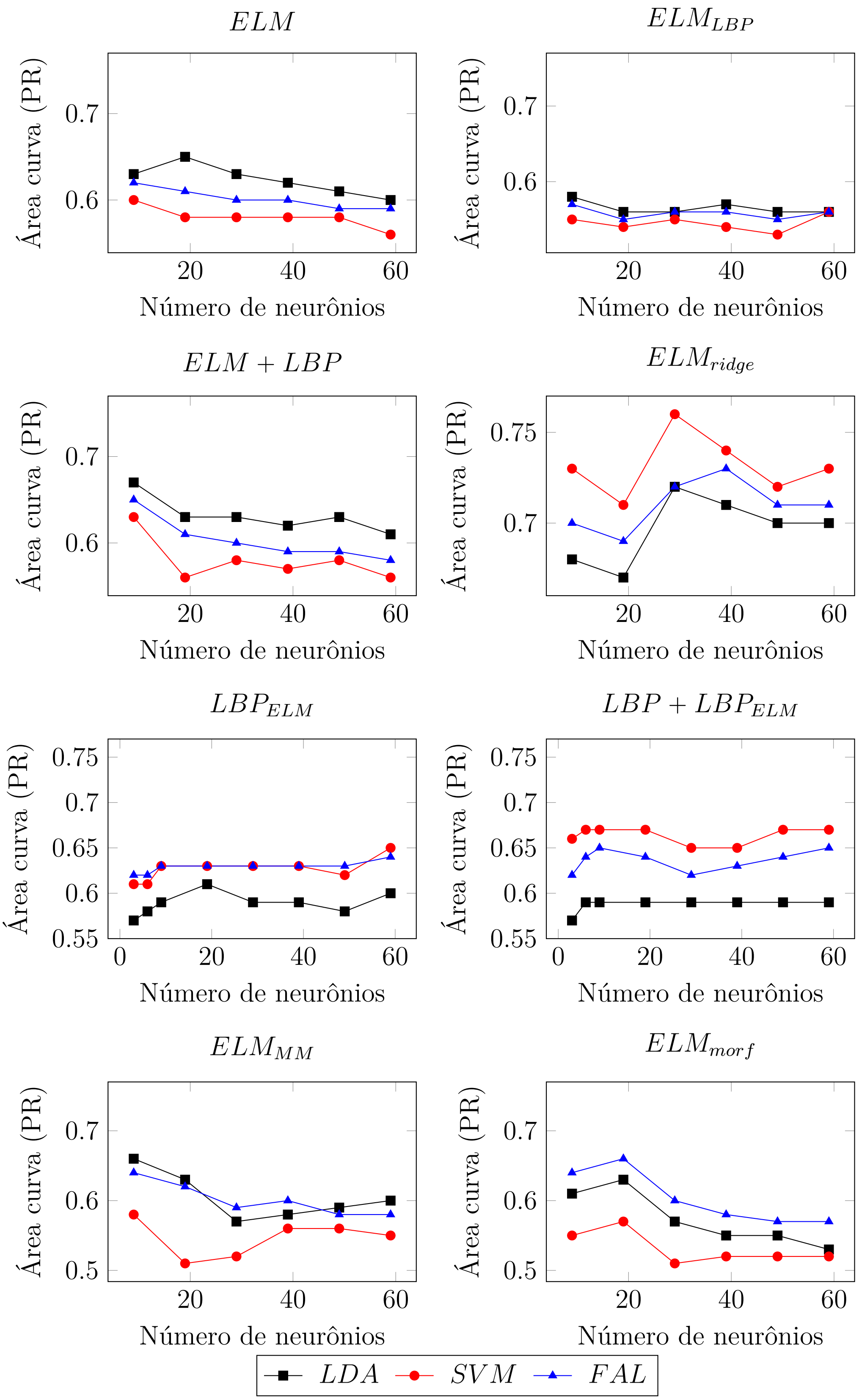


Figura 30 – Classificação de imagens com câncer × sem câncer.

que neste caso, como já havia ocorrido nas bases de *benchmark*, houve um ganho trazido por essas combinações, porém ainda não o suficiente para superar os descritores *ridge* individuais.

Tabela 8 – Combinação de diferentes métodos para o problema de duas classes: câncer × sem câncer.

	Área da curva PR		
Configuração	LDA	SVM	FAL
$(ELM_{ridge})_9 + (ELM_{ridge})_{19}$	0.73	0,66	0.72
$(ELM_{ridge})_9 + (ELM_{ridge})_{19} + (ELM_{ridge})_{29}$	0.75	0.76	0.73
$(ELM)_9 + (ELM + LBP)_9 + (ELM_{ridge})_9$	0.72	0.65	0.71
$(ELM)_{19} + (ELM + LBP)_{19} + (ELM_{ridge})_{19}$	0.72	0.65	0.69
$(ELM)_9 + (ELM_{LBP})_9 +$ $+ (ELM_{ridge})_9 + (ELM_{LBP})_9$	0.73	0.65	0.71
$(ELM)_{19} + (ELM_{LBP})_{19} +$ $+ (ELM_{ridge})_{19} + (ELM_{LBP})_{19}$	0.71	0.56	0.67
$(ELM)_9 + (ELM_{ridge})_9 +$ $+ (ELM_{MM})_9 + (ELM_{morf})_9 + (ELM + LBP)_9$	0.73	0.65	0.72

Como já vimos nas bases de *benchmark*, além da taxa de acerto e da área sob a curva PR, uma outra maneira de se avaliar a efetividade de um classificador é por meio da matriz de confusão. Aqui essa matriz é exibida na Figura 31 para a melhor configuração de classificação, isto é, $(ELM_{ridge})_9 + (ELM_{ridge})_{19} + (ELM_{ridge})_{29}$.

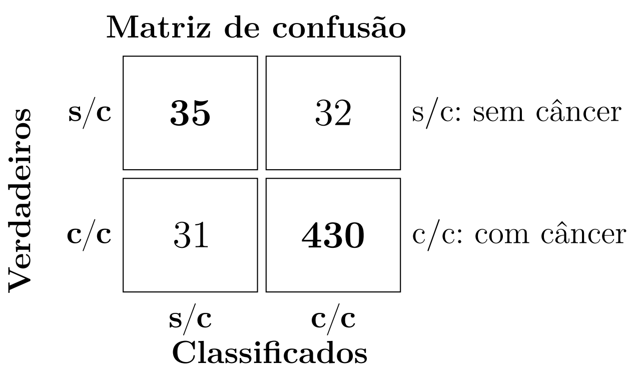


Figura 31 – Matriz de Confusão para a configuração $(ELM_{ridge})_9 + (ELM_{ridge})_{19} + (ELM_{ridge})_{29}$, no problema câncer × sem câncer.

A separação de 90 % das imagens para treinamento e o restante para teste nesta base implicou neste caso em 461 amostras com câncer e 67 sem câncer. Analisando-se a Figura 31, observa-se que em torno de 93% das imagens com câncer foram classificadas corretamente, enquanto que a detecção de imagens sem a doença apresentou aproximadamente 52% de acertos. Aqui é importante que se enfatize o fato de o classificador apresentar um baixo índice de falsos negativos, já que neste cenário isso implicaria em atraso da aplicação de um tratamento cuja urgência é fundamental.

Ainda na mesma configuração usada na matriz de confusão, foi avaliada também a curva de *precision/recall* (PR) como um todo, não apenas sua área. Como vimos, essa é outra forma de analisar os dados classificados, especialmente importante em situações em que ocorre um desbalanceamento no número de amostras em cada classe, como é o caso aqui.

Como vimos também, o valor da área sob esta curva é uma medida entre 0 e 1 que se assemelha sob vários aspectos ao conceito de acurácia (ou taxa de acerto) da classificação, sendo que valores tendendo a 1 indicam um bom desempenho na classificação dos dados. Para este problema, foi encontrado um valor de área igual a 0.76, como apresentado na Figura 32. No geral, a curva reflete um classificador com desempenho relativamente bom quando colocado em perspectiva com o desafio do problema, recuperando adequadamente as amostras de cada classe (*recall*) e classificando corretamente as imagens recuperadas (*precision*) na maioria dos casos.

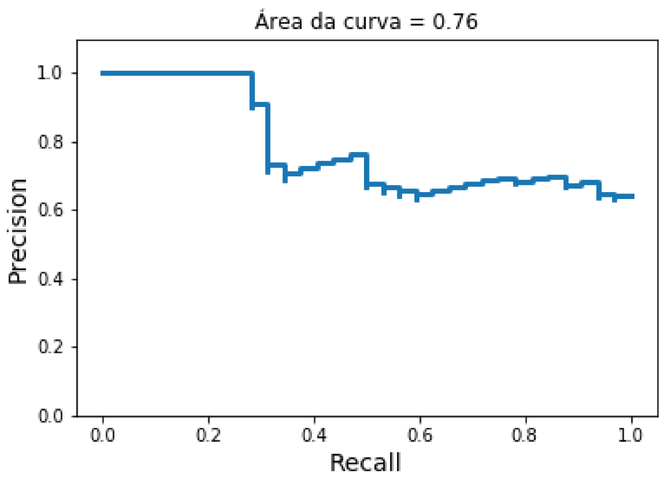


Figura 32 – Curva de *precision/recall* para a configuração $(ELM_{ridge})_9 + (ELM_{ridge})_{19} + (ELM_{ridge})_{29}$, câncer $\times$ sem câncer.

8.3.2 Pequenas Células *versus* Não Pequenas Células

Para o segundo teste de classificação nesta base de dados, as imagens também foram separadas em duas classes: câncer de pequenas células *versus* câncer de não pequenas células.

Na Figura 34 temos os resultados para as variações no número de neurônios da camada oculta para as metodologias baseadas no ELM. Para o método de regressão *ridge*, foi adotado o valor de $\gamma = 10^{-1}$ com base na Figura 33.

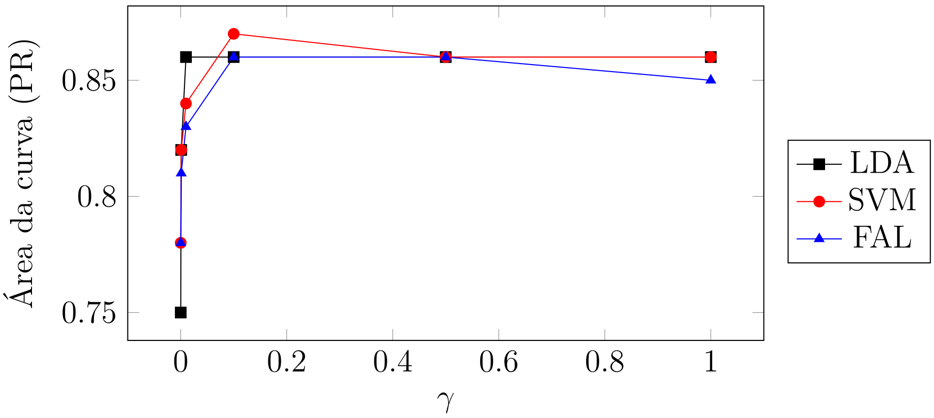


Figura 33 – Variação de γ $(10^{-4}, 10^{-3}, 10^{-2}, 10^{-1}, 0.5, 1)$ para o problema pequenas células *versus* não pequenas células.

Neste problema específico, nota-se pela Figura 34 que o ELM_{ridge} e o LBP_{ELM} se destacaram perante os demais. Assim como no primeiro problema, apenas 9 neurônios na camada oculta já foram suficientes para que se obtivesse uma taxa de acerto alta. O uso de mais neurônios, porém, teve maior impacto positivo neste cenário, como esperado de um problema de maior complexidade.

Já em relação às variações da transformação ELM em combinação com descritores LBP, os resultados foram superiores aos encontrados para a maioria das demais metodologias baseadas em ELM, sendo que a combinação com os descritores LBP sobre a imagem original propiciaram ainda um pequeno ganho na área da curva PR.

Como ocorrido nas bases de *benchmark*, houve aqui uma leve melhora na taxa de acerto geral com a concatenação de descritores, como atestado pelos resultados na Tabela 9.

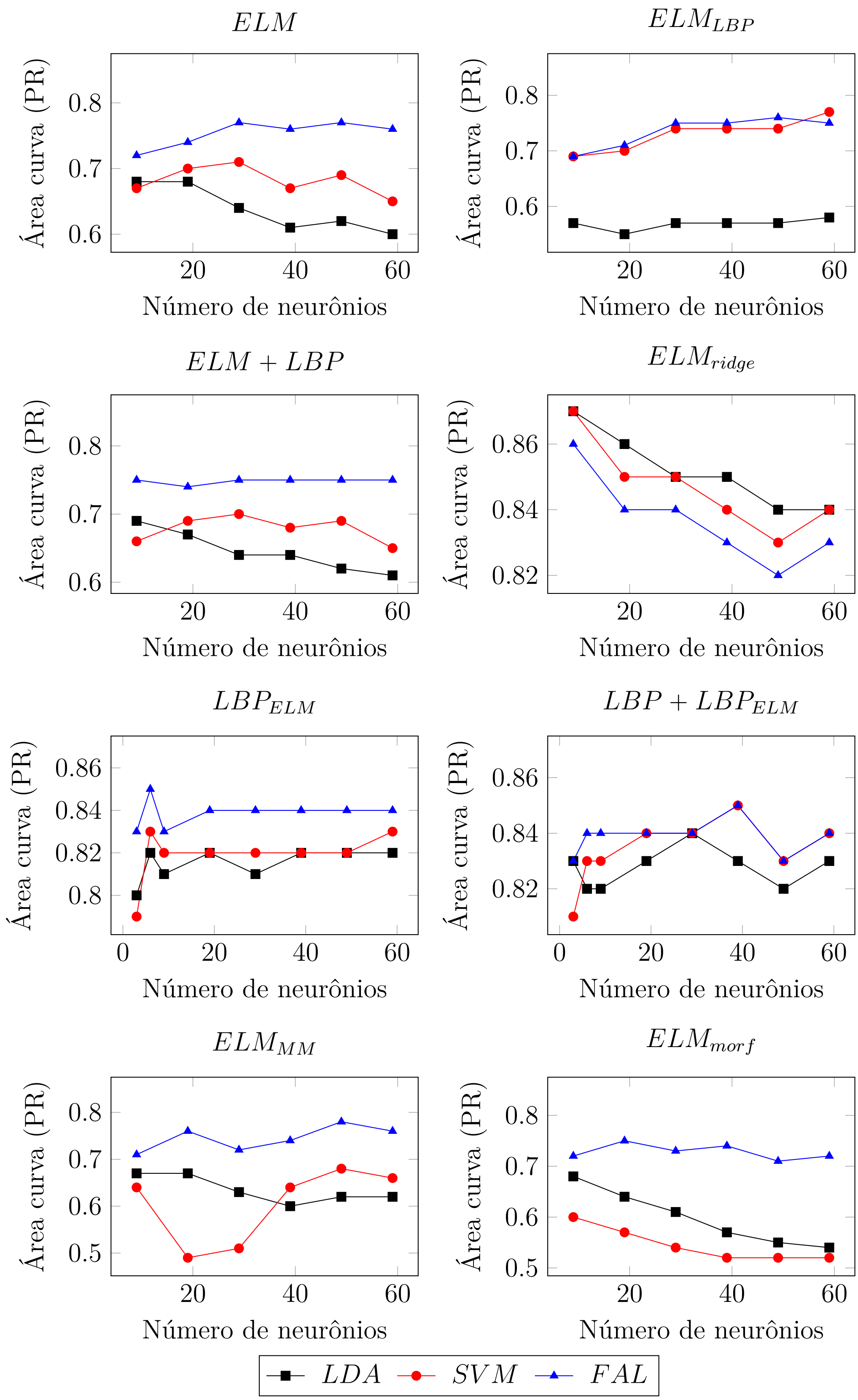


Figura 34 – Variação no número de neurônios para o problema câncer de pequenas células × não pequenas células.

Tabela 9 – Combinação de descritores baseados em ELM para o problema pequenas células *versus* não pequenas células.

	Área da curva PR		
Configuração	LDA	SVM	FAL
$(ELM + LBP)_9 + (ELM_{ridge})_9$	0.87	0.78	0.86
$(ELM)_9 + (ELM + LBP)_9 + (ELM_{ridge})_9$	0.87	0.77	0.86
$(ELM)_{19} + (ELM + LBP)_{19} + (ELM_{ridge})_{19}$	0.86	0.67	0.83
$(ELM)_{29} + (ELM + LBP)_{29} + (ELM_{ridge})_{29}$	0.87	0.73	0.85
$(ELM)_9 + ELM_{ridge})_9 +$ $+(ELM_{MM})_9 + (ELM_{morf})_9$	0.87	0.76	0.87
$(ELM)_{19} + (ELM + LBP)_{19} +$ $+(ELM_{ridge})_{19} + (ELM_{LBP})_{19}$	0.87	0.78	0.86
$(ELM_{ridge})_9 + (ELM_{ridge})_{19} + (ELM_{ridge})_{29}$	0.88	0.87	0.86

Semelhantemente ao teste anterior, também foi gerada aqui a matriz de confusão exibida na Figura 35 com o resultado para a configuração $(ELM_{ridge})_9 + (ELM_{ridge})_{19} + (ELM_{ridge})_{29}$.

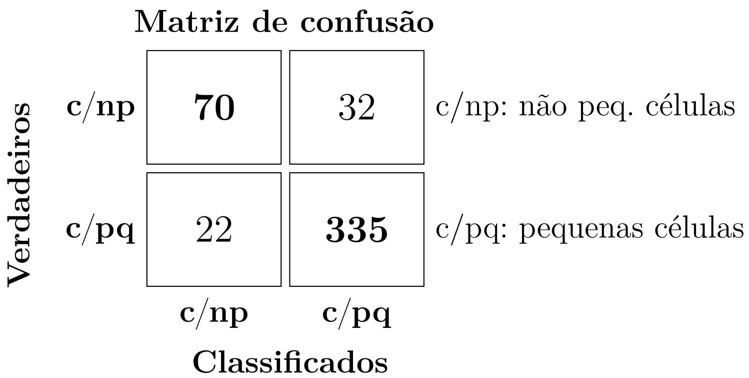


Figura 35 – Matriz de confusão para o problema câncer de não pequenas células *versus* câncer de pequenas células.

Observando-se a Figura 35, nota-se em torno de 69 % de acerto na identificação de imagens com câncer de não pequenas células e 94 % para pequenas células. Esta melhora na classificação em relação ao problema anterior também pode ser observada na área da curva PR igual 0.88, como destacado na Figura 36. Em geral, o desempenho da metodologia apresentada neste problema é bastante relevante do ponto de vista prático, já que, como visto, a identificação do sub-tipo correto tem importância crucial tanto do ponto de vista do diagnóstico em si quanto do prognóstico.

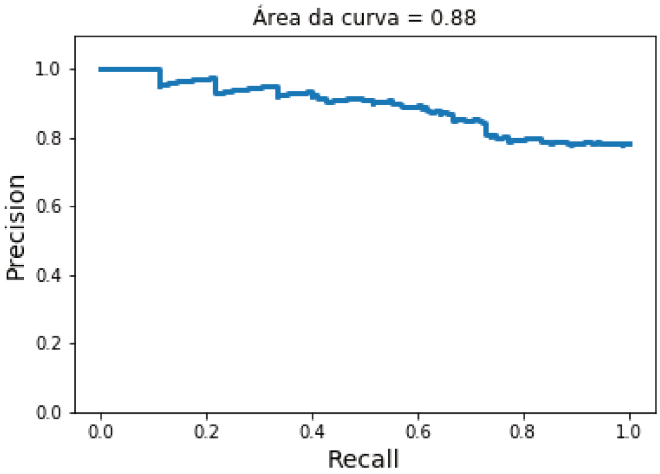


Figura 36 – Curva de *precision/recall* para o problema de câncer de não pequenas células *versus* câncer de pequenas células .

8.3.3 Adenocarcinoma *versus* Epidermoide

O terceiro teste de classificação de imagens médicas ainda considera duas classes, porém agora ambas do tipo de câncer de não pequenas células, dos sub-tipos adenocarcinoma e epidermoide. Para o método com regressão *ridge*, foi utilizado $\gamma = 1$, com base na Figura 37. Na Figura 38 temos os resultados para a variação no número de neurônios da camada oculta.

Observando-se a Figura 38, nota-se que em várias das metodologias comparadas ainda houve uma área PR próxima da máxima com 9 neurônios na camada oculta, como anteriormente. Observa-se, de fato, que para algumas das variantes testadas o desempenho chegou até mesmo a ser pior com o aumento de neurônios, o que demonstra o grau de dificuldade do problema e que o uso de mais neurônios apenas enfatizou um *overfitting* intrínseco ao modelo. Grande parte dessa dificuldade advém da diferença nas texturas nos dois sub-tipos de não pequenas células, que é bem mais sutil do que nas configurações anteriores.

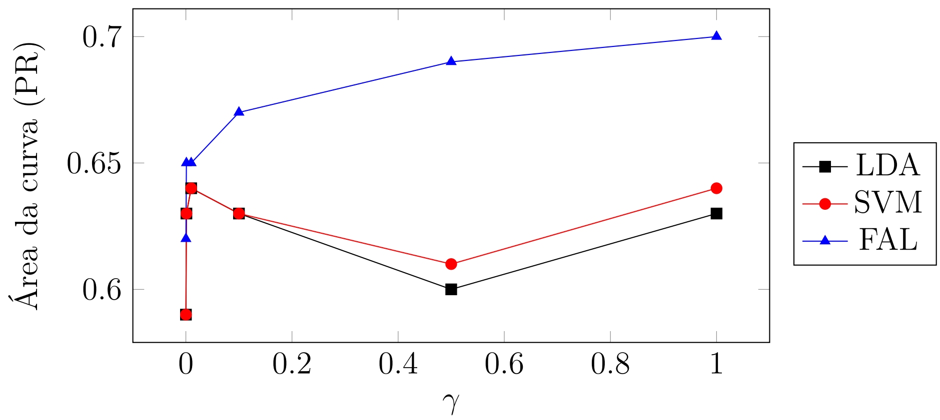


Figura 37 – Variação de γ $(10^{-4}, 10^{-3}, 10^{-2}, 10^{-1}, 0.5, 1)$ para o problema adenocarcinoma *versus* epidermoide.

O desempenho geral é similar ao obtido no primeiro problema médico, câncer *vs* normal e o método ELM_{ridge} teve novamente o melhor desempenho entre as variantes comparadas. Os resultados com variações da transformação ELM, LBP_{ELM}, também seguiram o mesmo padrão observado até o momento. Novamente, os descritores LBP sobre a imagem original não trouxeram ganho substancial ao desempenho do classificador.

A concatenação de descritores não trouxe ganho significativo para a área sob a curva PR, como ilustrado na Tabela 10. Aqui temos um problema mais complexo do que os anteriores e este é de fato um caso em que mesmo a combinação de atributos potencialmente complementares não foi capaz de contribuir para um desempenho mais consistente e robusto.

Tanto a matriz de confusão na Figura 39 quanto a curva PR na Figura 40 foram calculadas para a melhor configuração de classificação $(ELM_{ridge})_9 + (ELM_{ridge})_{19} + (ELM_{ridge})_{29}$.

De acordo com resultados apresentados na Figura 39, em torno de 67 % das imagens da classe adenocarcinoma foram classificadas corretamente e 62 % das do subtipo epidermoide foram identificadas, correspondendo também a um valor de 0.72 da área sob a curva PR na Figura 40. Ambos os resultados mostram um classificador equilibrado em números de falsos positivos e falsos negativos e com desempenho relativamente bom se considerada a dificuldade inerente ao problema.

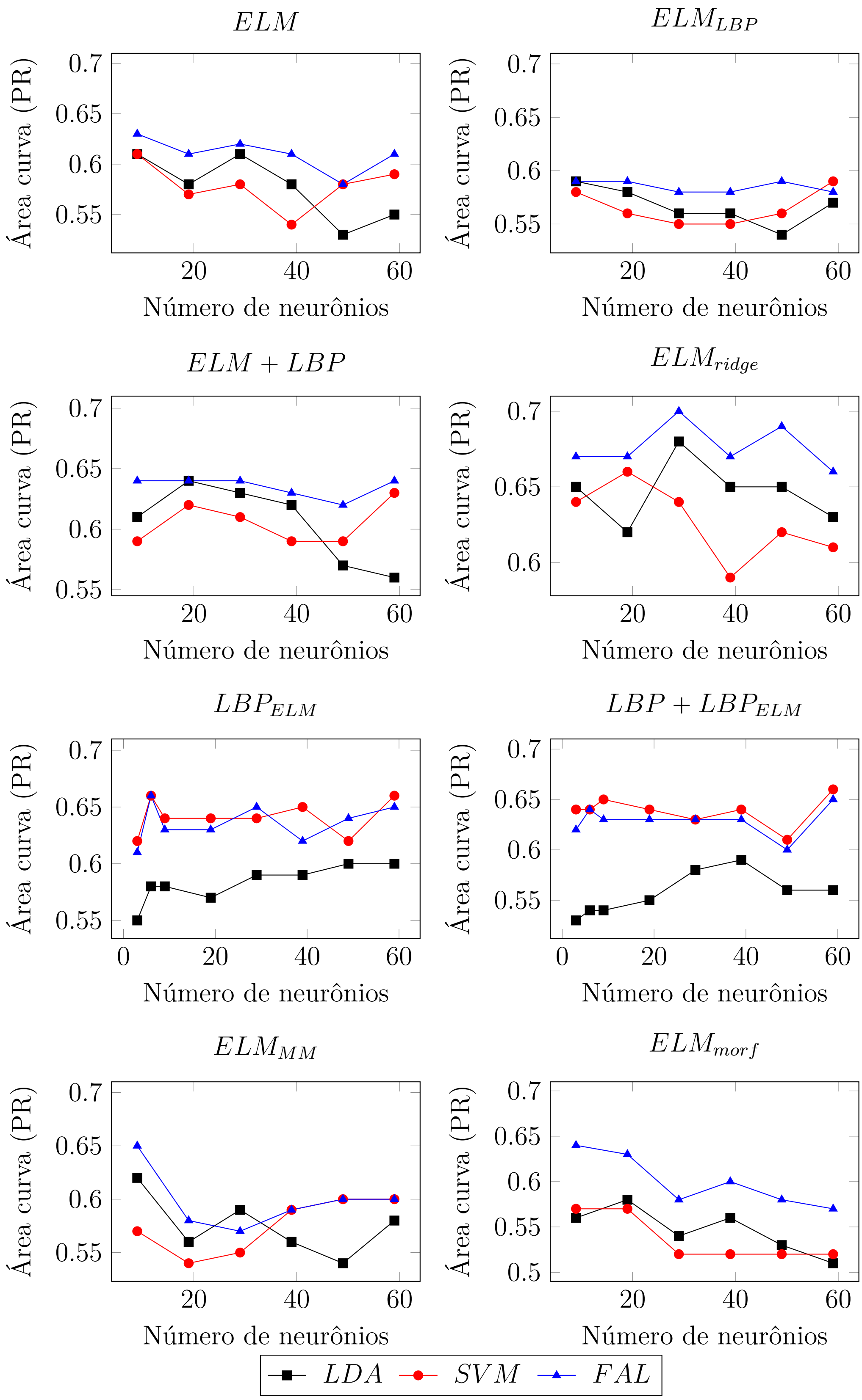


Figura 38 – Classificação do problema adenocarcinoma $\times$ epidermoide variando-se o número de neurônios na camada oculta.

Tabela 10 – Combinações de descritores ELM para o problema adenocarcinoma *versus* epidermoide.

	Área da curva PR		
Configuração	LDA	SVM	FAL
$(ELM + LBP)_9 + (ELM_{ridge})_9$	0.65	0.59	0.70
$(ELM + LBP)_{19} + (ELM_{ridge})_{19}$	0.68	0.60	0.70
$(ELM + LBP)_{29} + (ELM_{ridge})_{29}$	0.68	0.60	0.66
$(ELM)_{29} + (ELM + LBP)_{29} + (ELM_{ridge})_{29}$	0.67	0.57	0.66
$(ELM)_9 + (ELM + LBP)_9 +$ $+ (ELM_{ridge})_9 + (ELM_{LBP})_9$	0.67	0.59	0.69
$(ELM)_9 + (ELM)_{19} + (ELM)_{29}$	0.61	0.60	0.64
$(ELM + LBP)_9 + (ELM + LBP)_{19} +$ $+ (ELM + LBP)_{29}$	0.67	0.64	0.67
$(ELM_{ridge})_9 + (ELM_{ridge})_{19} + (ELM_{ridge})_{29}$	0.71	0.64	0.72

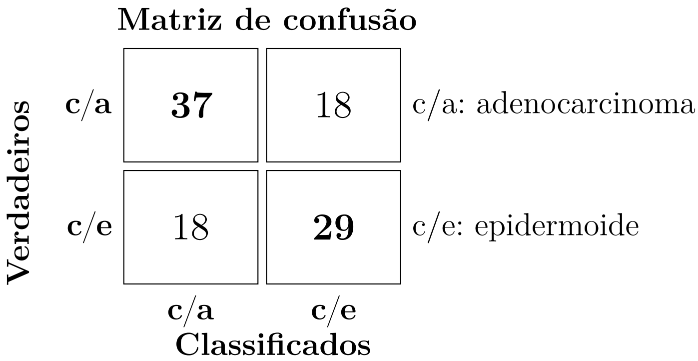


Figura 39 – Classes: adenocarcinoma × epidermoide.

8.3.4 Normal *versus* Adenocarcinoma *versus* Epidermoide *versus* Pequenas Células

Por fim, o último teste de classificação envolveu todas as quatro classes da base de dados: normal (sem câncer), adenocarcinoma, epidermoide e pequenas células.

Para o ELM com regressão *ridge*, similar ao que foi feito nas bases do capítulo 7, variou-se o valor de γ para 9 neurônios na camada oculta e os resultados estão na Figura 41.

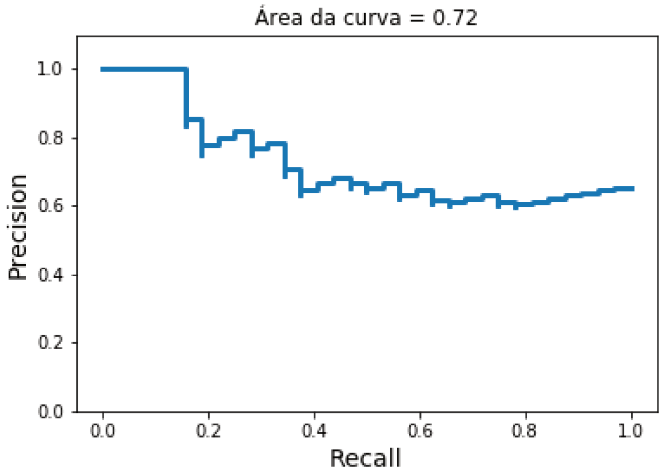


Figura 40 – Curva de *precision/recall* para o problema adenocarcinoma *versus* epidermoide.

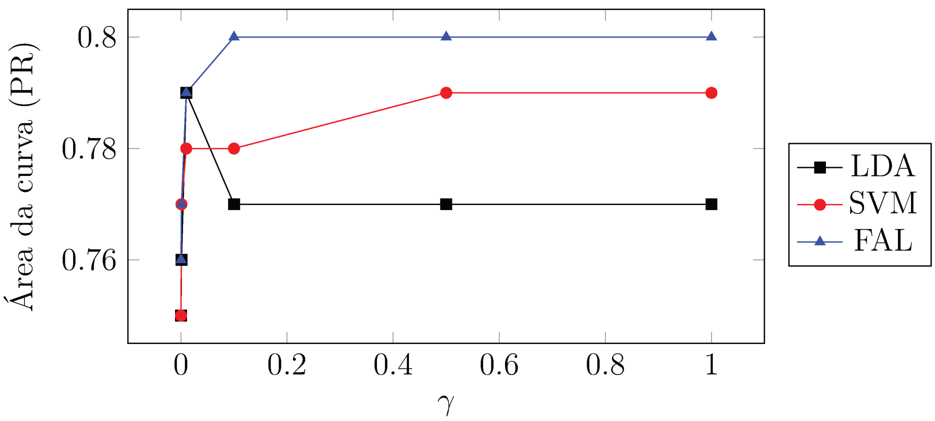


Figura 41 – Variação de γ $(10^{-4}, 10^{-3}, 10^{-2}, 10^{-1}, 0.5, 1)$ para o problema normal *versus* adenocarcinoma *versus* epidermoide *versus* pequenas células.

De acordo com os resultados encontrados na Figura 41, adotou-se o valor de $\gamma = 10^{-1}$ para o algoritmo com regressão *ridge*. Este e outros resultados são apresentados então na Figura 42.

Os resultados foram semelhantes, porém a única configuração que apresentou área sob a curva PR superior a 0.8 foi o ELM com regressão *ridge*.

Além das configurações da Figura 42, testes foram feitos para a concatenação de descritores ELM e os resultados estão na Tabela 11. Novamente não houve variação sensível na taxa de acerto com as concatenações.

Como antes, foi avaliada também a matriz de confusão para a melhor configuração, que neste caso foi $(ELM)_9 + (ELM + LBP)_9 + (ELM_{ridge})_9$, como apresentado

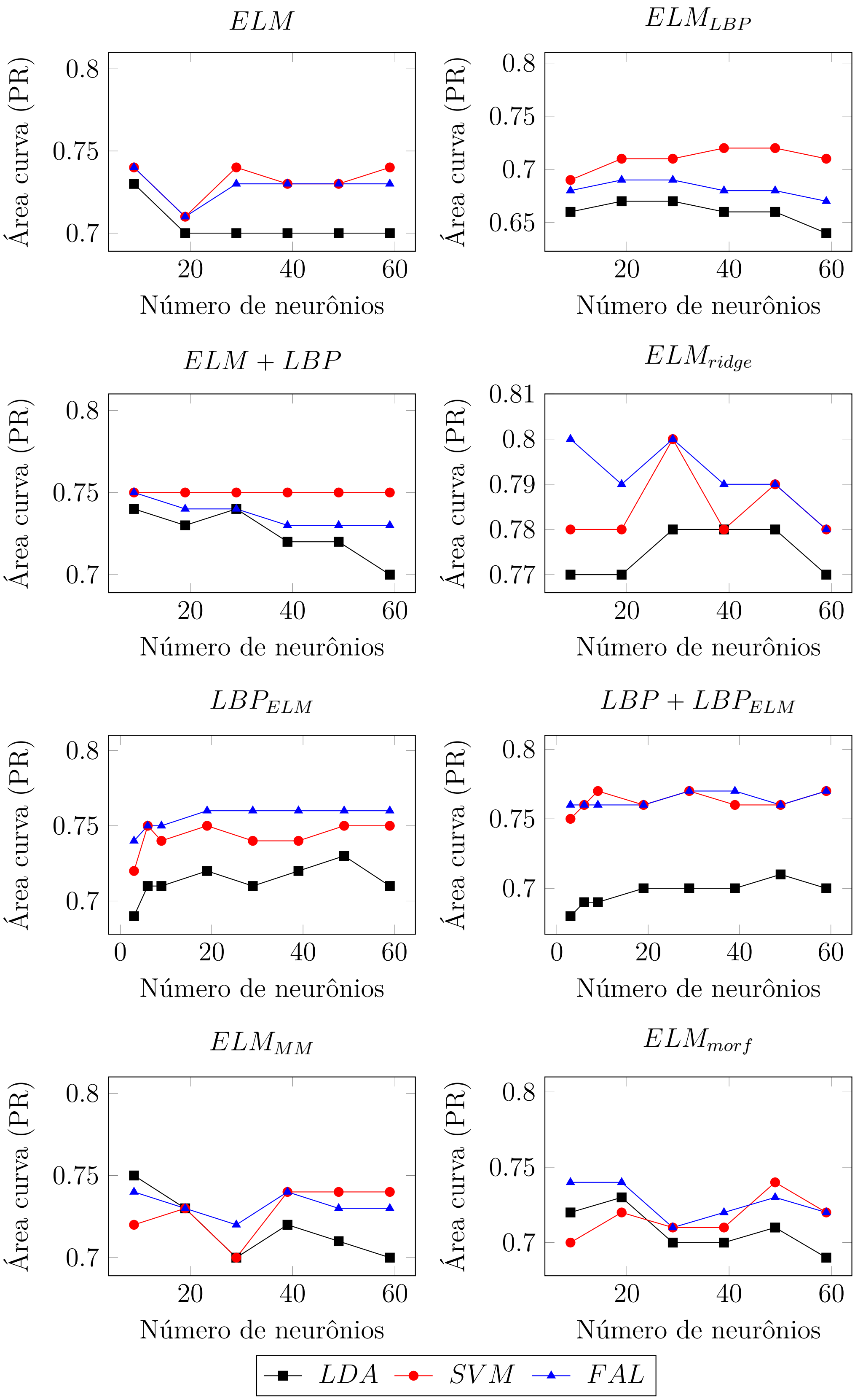


Figura 42 – Variação no número de neurônios para o problema médico de quatro classes.

Tabela 11 – Combinação de descritores ELM no problema médico de quatro classes.

	Área da curva PR		
Configuração	LDA	SVM	FAL
$(ELM)_9 + (ELM + LBP)_9 + (ELM_{ridge})_9$	0.79	0.81	0.81
$(ELM)_{19} + (ELM + LBP)_{19} + (ELM_{ridge})_{19}$	0.77	0.79	0.78
$(ELM)_9 + (ELM + LBP)_9 +$ $+(ELM_{ridge})_9 + ELM_{LBP_9}$	0.79	0.81	0.80
$(ELM)_{19} + (ELM + LBP)_{19} +$ $+(ELM_{ridge})_{19} + (ELM_{LBP})_{19}$	0.78	0.80	0.79

na Figura 43.

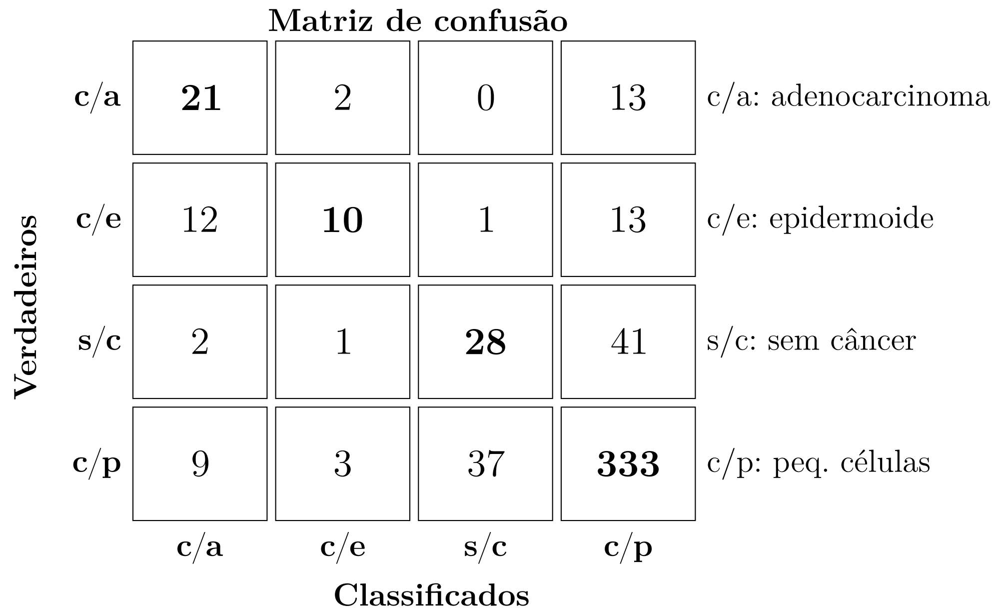


Figura 43 – Matriz de confusão para o problema médico de quatro classes.

Neste caso, a matriz de confusão reflete o cenário de um conjunto de teste com 526 imagens, sendo divididas em 36 de câncer adenocarcinoma, 36 epidermoide, 72 sem câncer e 382 com câncer de pequenas células. Foram classificadas corretamente em torno de 58% das imagens com câncer adenocarcinoma, 28% das com câncer epidermoide, 39% sem câncer e 87% das amostras com câncer de pequenas células. Vale ressaltar também que 80% dos casos de câncer (nos 3 sub-tipos) foram identificados corretamente.

A matriz de confusão é complementada pela curva de *precision/recall* na Figura 44, a qual exibe também uma área sob a curva igual a 0.82, desempenho interessante quando se considera o número maior de classes e o desafio inerente ao uso de imagens de núcleos celulares avulsos nesta base.

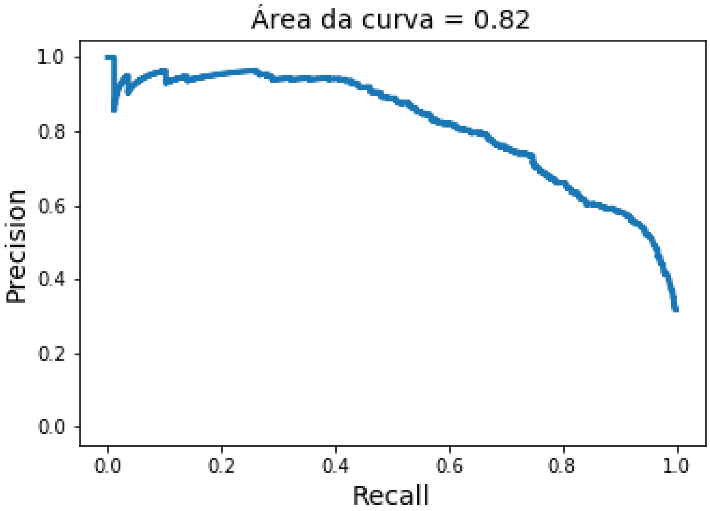


Figura 44 – Curva de *precision/recall* para o problema médico de quatro classes.

8.4 Observações Finais

Entre os desafios deste problema das imagens de câncer de pulmão, há o desbalanceamento no número de imagens para cada classe, prevalecendo um número superior de amostras com a presença do câncer de pequenas células. Consequentemente, acaba havendo um peso maior para o desempenho do classificador nesta classe.

Além de diferentes números de imagens para cada classe, esta base também tem a característica de as imagens variarem de tamanho, sendo este sempre inferior a 256×256 a partir do momento em que o fundo branco em volta da região elipsoidal do núcleo é descartado.

De um modo geral, as áreas sob a curva PR das metodologias testadas para esta base ficaram próximas entre si, com leve destaque para a regressão *ridge*, assim como também ocorreu nas bases de *benchmark*.

Outra observação interessante referente a esta base é que os resultados da abordagem com aplicação do ELM em imagens após a transformada LBP alcançou resultados mais próximos do ELM padrão, diferentemente do que ocorreu na bases de *benchmark*. O desafio intrínseco ao problema médico aqui apresentado é uma potencial razão para este fenômeno, fazendo com que a diferença de desempenho entre os descritores seja menos marcante.

Os melhores resultados de cada método abordado neste trabalho e para cada problema de separação entre imagens médicas abordado neste capítulo são apresentados na Tabela 12.

Para comparação das metodologias abordadas neste estudo para a extração de descritores, foram testados os seguintes métodos: dimensão fractal de Bouligand-Minkowski (BACKES et al., 2009), transformada de Fourier (AZENCOTT et al., 1997), matriz de co-ocorrência de níveis de cinza (em inglês *Gray Level Co-occurrence Matriz* - GLCM) (HARALICK et al., 1973), filtro de Gabor (MANJUNATH; MA, 1996), padrões locais binários (LBP) (Ojala et al., 2002) e Redes Neurais Convolucionais (sigla em inglês - CNN) (CIMPOI et al., 2016).

Os resultados destes procedimentos são apresentados na Tabela 13, destacando para cada método os valores da área da curva PR para os classificadores LDA, SVM e FAL.

Para avaliar os resultados apresentados na Tabela 12, destacaremos as taxas de acertos das novas abordagens para extrair descritores de imagens neste trabalho: ELM_{LBP}, $ELM + LBP$ e LBP_{ELM}, os melhores valores das metologias já existentes ELM, ELM_{ridge}, ELM_{MM} e ELM_{morf} para comparar com os métodos da Tabela 13.

- Câncer × sem câncer: nos métodos apresentados neste trabalho houve destaque para $ELM + LBP$ e ELM_{ridge}, ambos ficaram superior as abordagens de Minkowski, Fourier, GLSM, Gabor e LBP, sendo inferior apenas para CNN;

- Pequenas células × não pequenas células: LBP_{ELM} foi superior a três métodos e com valores iguais a Gabor e LBP, ficando com menor taxa de acerto quando comparado com CNN, enquanto ELM_{ridge} é superior a cinco abordagens,

Tabela 12 – Maiores taxas de acerto para cada método e cada problema de imagens médicas aqui considerado.

Configuração	Área (PR)	Configuração	Área (PR)
Câncer × sem câncer			
$(ELM)_{19}$	0.65	$(LBP_{ELM})_{59}$	0.65
$(ELM_{LBP})_{9}$	0.58	$LBP + (LBP_{ELM})_{6}$	0.67
$(ELM + LBP)_{9}$	0.67	$(ELM_{MM})_{9}$	0.66
$(ELM_{ridge})_{39}$	0.74	$(ELM_{morf})_{19}$	0.66
$(ELM_{ridge})_{9} + (ELM_{ridge})_{19} + (ELM_{ridge})_{29}$			0.76

Pequenas células × não pequenas células			
$(ELM)_{29}$	0.77	$(LBP_{ELM})_{6}$	0.85
$(ELM_{LBP})_{59}$	0.77	$LBP + (LBP_{ELM})_{39}$	0.85
$(ELM + LBP)_{9}$	0.75	$(ELM_{MM})_{49}$	0.78
$(ELM_{ridge})_{9}$	0.87	$(ELM_{morf})_{19}$	0.75
$(ELM_{ridge})_{9} + (ELM_{ridge})_{19} + (ELM_{ridge})_{29}$			0.88

Adenocarcinoma × epidermoide			
$(ELM)_{9}$	0.63	$(LBP_{ELM})_{6}$	0.66
$(ELM_{LBP})_{9}$	0.59	$LBP + (LBP_{ELM})_{59}$	0.66
$(ELM + LBP)_{9}$	0.64	$(ELM_{MM})_{9}$	0.65
$(ELM_{ridge})_{29}$	0.70	$(ELM_{morf})_{9}$	0.64
$(ELM_{ridge})_{9} + (ELM_{ridge})_{19} + (ELM_{ridge})_{29}$			0.72

Adenocarcinoma × epidermoide × pequenas células × sem câncer			
$(ELM)_{9}$	0.74	$(LBP_{ELM})_{19}$	0.76
$(ELM_{LBP})_{39}$	0.72	$LBP + (LBP_{ELM})_{9}$	0.77
$(ELM + LBP)_{9}$	0.75	$(ELM_{MM})_{9}$	0.75
$(ELM_{ridge})_{9}$	0.80	$(ELM_{morf})_{9}$	0.74
$ELM)_{9} + (ELM + LBP)_{9} + (ELM_{ridge})_{9}$			0.81

Tabela 13 – Outros métodos para extração de descritores na base de imagens médicas e classificação.

	Área da curva PR		
Método	LDA	SVM	FAL
Câncer × sem câncer			
Minkowski	0.64	0.53	0.57
Fourier	0.61	0.61	0.60
GLSM	0.61	0.53	0.56
Gabor	0.64	0.55	0.63
LBP	0.58	0.58	0.61
CNN	0.68	0.81	0.78
Pequenas células × não pequenas células			
Minkowski	0.84	0.77	0.77
Fourier	0.63	0.81	0.78
GLSM	0.71	0.69	0.63
Gabor	0.85	0.80	0.84
LBP	0.83	0.84	0.85
CNN	0.59	0.92	0.90
Adenocarcinoma × epidermoide			
Minkowski	0.63	0.64	0.62
Fourier	0.53	0.60	0.60
GLSM	0.63	0.62	0.59
Gabor	0.69	0.63	0.68
LBP	0.54	0.64	0.63
CNN	0.66	0.72	0.73
Adenocarcinoma × epidermoide × pequenas células × sem câncer			
Minkowski	0.70	0.67	0.71
Fourier	0.70	0.73	0.71
GLSM	0.71	0.67	0.67
Gabor	0.72	0.73	0.76
LBP	0.71	0.75	0.75
CNN	0.40	0.86	0.84

ficando menor apenas para CNN;

- Adenocarcinoma × epidermoide: LBP_{ELM} foi superior a Minkowski, Fourier, GLCM e LBP, ficando com valores menores nos outros dois métodos, enquanto que o ELM_{ridge} ficou menor apenas para CNN;

- Adenocarcinoma × epidermoide × pequenas células × sem câncer: LBP_{ELM} apresentou melhor resultado que: Minkowski, Fourier, GLCM e LBP, com valor igual a Gabor e inferior a CNN. Novamente ELM_{ridge} ficou inferior apenas a CNN;

De acordo com as comparações realizadas anteriormente, as novas abordagens para extrair vetores de característica de imagens apresentaram-se eficientes também na base de imagens médicas, destacando-se o LBP_{ELM}, em que é aplicada a transformação ELM às imagens para extrair descritores.

A utilização de regressão *ridge* na resolução das equações do ELM também apresentou bons resultados, ficando superior a Minkowski, Fourier, GLCM, Gabor e LBP. Ficou com resultado inferior apenas em relação às redes neurais convolucionais, mas este procedimento apresenta custo computacional significativamente maior.

Como vimos nos testes das bases de *benchmark*, na base de imagens médicas os métodos abordados neste trabalho também apresentaram competitividade com as metodologias já existentes.

Em geral, os testes apresentaram resultados satisfatórios para a identificação da presença ou não do câncer de pulmão. Quando os sub-tipos são analisados, destaca-se a taxa de identificação do câncer de pequenas células. Como visto no início deste capítulo, este é um resultado bastante interessante do ponto de vista prático, já que este é exatamente o sub-tipo para o qual um diagnóstico o mais precoce possível é crucial para um tratamento bem sucedido.

9 Considerações Finais

Neste trabalho foram apresentadas variações de um algoritmo baseado em máquinas de aprendizado extremo para extrair descritores locais de texturas (SÁ JUNIOR; BACKES, 2016). Entre as abordagens aqui desenvolvidas, tivemos três novas propostas, baseadas em uma combinação no nível de descritores de ELM com LBP, aplicação do ELM em transformadas LBP, LBP aplicado sobre uma transformação ELM, transformada esta que foi apresentada pela primeira vez neste trabalho. Ainda foi trabalhado o ELM com regressão *ridge* e ELM morfológico.

Para avaliar esses métodos, testes de desempenho foram desenvolvidos em quatro bases de imagens, sendo três de *benchmark* e uma voltada para um problema do mundo real na área médica.

A combinação direta por meio de concatenação de descritores ELM com LBP apresentou bons resultados nas bases de *benchmark*, principalmente para nove neurônios na camada oculta. Resultados promissores também foram observados nessas bases para a utilização do ELM com regressão *ridge*.

Já a aplicação do ELM em imagens após a transformada LBP não apresentou melhora significativa nas bases de *benchmark*, mas no problema de classificação de câncer de pulmão alcançou resultados competitivos com o estado-da-arte.

Por sua vez, o uso da transformação ELM aqui apresentado, seguido da extração de descritores LBP, trouxe acréscimos significativos, especialmente quando descritores da imagem original foram combinados com descritores da transformada ELM.

Outro fato observado nos testes executados é o efeito da variação no classificador utilizado. Por exemplo, na Figura 23 o LDA alcançou melhores resultados, enquanto que na Figura 26 o FAL se sobressaiu. Na base de dados de imagens médicas, os três classificadores também apresentaram comportamento semelhante em boa parte dos testes.

Para a base de imagens com diferentes tipos de câncer de pulmão, a maior classificação correta foi nos testes que visavam discriminar o sub-tipo de pequenas células do não pequenas células, o que é um resultado bastante interessante considerando-se a importância do diagnóstico precoce do câncer de pequenas células. Quando o número de classes foi aumentado, a taxa de acerto caiu, fato esperado pela maior dificuldade trazida pelo aumento no número de classes.

A metodologia para classificação de imagens médicas apresentada neste estudo pode servir de apoio para patologistas na identificação da presença ou não do câncer de pulmão e de seu sub-tipo, servindo assim como uma segunda opinião ou uma primeira avaliação imediata após a disponibilização da imagem microscópica da amostra.

Para trabalhos futuros, pretende-se aprimorar os algoritmos aqui estudados, visando alcançar uma melhor generalização do modelo para o conjunto de testes, isso também aliado a um menor custo computacional. Conceitos matemáticos envolvidos nas modificações dos algoritmos aqui apresentados também serão estudados, visando assim uma análise mais formal e aprofundada da motivação teórica por trás de cada resultado aqui obtido.

Por fim, uma outra exploração possível para o futuro baseia-se em uma utilização mais efetiva da teoria da análise assintótica para a construção do algoritmo ELM, visando também um aperfeiçoamento tanto no tempo de execução quanto na precisão e robustez dos descritores para os testes de classificação.

Referências

ADAM, R. et al. The fractal dimension of nuclear chromatin as a prognostic factor in acute precursor b lymphoblastic leukemia. *Cellular oncology : the official journal of the International Society for Cellular Oncology*, v. 28, p. 55–59, 02 2006. Disponível em: <https://doi.org/10.1155/2006/409593>. Citado na página 72.

AZENCOTT, R.; WANG, J.-P.; YOUNES, L. Texture classification using windowed fourier filters. *IEEE Trans. Pattern Anal. Mach. Intell.*, v. 19, n. 2, p. 148–153, 1997. Disponível em: <http://dblp.uni-trier.de/db/journals/pami/pami19.html#AzencottWY97>. Citado na página 89.

BACKES, A.; CASANOVA, D.; BRUNO, O. Plant leaf identification based on volumetric fractal dimension. *IJPRAI*, v. 23, p. 1145–1160, 09 2009. Citado na página 89.

BACKES, A. et al. Randomized neural network based signature for color texture classification. *Multidimensional Systems and Signal Processing*, Springer, Van Godewijckstraat 30, 3311 GZ Dordrecht, Netherlands, 30, n. 3, p. 1171–1186, JUL 2019. ISSN 0923-6082. Disponível em: <https://doi.org/10.1007/s11045-018-0600-6>. Citado 2 vezes nas páginas 17 e 38.

BRADSKI, G. The OpenCV Library. *Dr. Dobb's Journal of Software Tools*, 2000. Citado na página 36.

BREIMAN, L. Random forests. *Machine Learning*, v. 45, n. 1, p. 5–32, Oct 2001. ISSN 1573-0565. Disponível em: <https://doi.org/10.1023/A:1010933404324>. Citado na página 11.

BRUNA, J.; MALLAT, S. Invariant scattering convolution networks. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, v. 35, n. 8, p. 1872–1886, 2013. Disponível em: <https://doi.org/10.1109/TPAMI.2012.230>. Citado na página 67.

CALVETTI, D. et al. Tikhonov regularization and the l-curve for large discrete ill-posed problems. *Journal of Computational and Applied Mathematics*, v. 123, n. 1, p. 423 – 446, 2000. ISSN 0377-0427. Numerical Analysis 2000. Vol. III: Linear Algebra. Disponível em: <http://www.sciencedirect.com/science/article/pii/S0377042700004143>. Citado na página 38.

CHAN, T. et al. PCANet: A simple deep learning baseline for image classification? *IEEE Transactions on Image Processing*, v. 24, n. 12, p. 5017–5032, 2015. Disponível em: <https://doi.org/10.1109/TIP.2015.2475625>. Citado na página 67.

CIMPOI, M. et al. Describing textures in the wild. In: *Proceedings of the 2014 IEEE Conference on Computer Vision and Pattern Recognition*. Washington, DC, USA: IEEE Computer Society, 2014. (CVPR '14), p. 3606–3613. Disponível em: <https://doi.org/10.1109/CVPR.2014.461>. Citado na página 67.

_____. Deep filter banks for texture recognition, description, and segmentation. *International Journal of Computer Vision*, v. 118, n. 1, p. 65–94, 2016. Disponível em: <https://doi.org/10.1007/s11263-015-0872-3>. Citado 2 vezes nas páginas 67 e 89.

GELENBE, E. Random neural networks with negative and positive signals and product form solution. *Neural Computation*, v. 1, n. 4, p. 502–510, 1989. Disponível em: <https://doi.org/10.1162/neco.1989.1.4.502>. Citado na página 40.

GOLUB, G. H.; LOAN, C. F. V. *Matrix Computations*. 4. ed. [S.l.]: Beltimore, 2013. Citado 2 vezes nas páginas 11 e 20.

GUO, Z.; ZHANG, L.; ZHANG, D. A completed modeling of local binary pattern operator for texture classification. *IEEE Transactions on Image Processing*, IEEE Press, v. 19, n. 6, p. 1657–1663, 2010. Disponível em: <https://doi.org/10.1109/TIP.2010.2044957>. Citado na página 67.

HARALICK, R.; SHANMUGAM, K.; DINSTEIN, I. Texture features for image classification. *IEEE Transactions on Systems, Man, and Cybernetics*, v. 3, n. 6, 1973. Citado na página 89.

HASTIE, T.; TIBSHIRANI, R.; FRIEDMAN, J. *The Elements of Statistical Learning: Data Mining, Inference, and prediction*. 1. ed. [S.l.]: New Yourk, 2001. Citado 3 vezes nas páginas 10, 11 e 53.

HAYKIN, S. *Redes neurais: princípios e prática*. [S.l.]: Porto Alegre, 2001. Citado na página 16.

HAYMAN, E. et al. On the significance of real-world conditions for material classification. In: PAJDLA, T.; MATAS, J. (Ed.). *Computer Vision - ECCV 2004*. Berlin, Heidelberg: Springer Berlin Heidelberg, 2004. p. 253–266. Disponível em: <https://doi.org/10.1007/978-3-540-24673-2_21>. Citado 3 vezes nas páginas 15, 47 e 48.

HUANG, G.-B.; ZHU, Q.-Y.; SIEW, C.-K. Extreme learning machine: Theory and applications. *Neurocomputing*, v. 70, n. 1, p. 489 – 501, 2006. ISSN 0925-2312. Neural Networks. Disponível em: <http://www.sciencedirect.com/science/article/pii/S0925231206000385>. Citado 10 vezes nas páginas 11, 12, 16, 18, 19, 20, 22, 24, 25 e 35.

HUANG, J. et al. Extreme learning machine with multi-scale local receptive fields for texture classification. *Multidimensional Systems and Signal Processing*, v. 28, n. 3, p. 995–1011, jul. 2017. ISSN 1573-0824. Disponível em: <https://doi.org/10.1007/s11045-016-0414-3>. Citado na página 18.

IDRISSA, M.; ACHEROY, M. Texture classification using gabor filters. *Pattern Recognition Letters*, v. 23, n. 9, p. 1095 – 1102, 2002. ISSN 0167-8655. Disponível em: <http://www.sciencedirect.com/science/article/pii/S0167865502000569>. Citado na página 28.

INCA. *Estatísticas de câncer*. 2020. Disponível em: <https://www.inca.gov.br/numeros-de-cancer>. Acesso em: 25 mar. 2020. Citado na página 39.

JUNIOR, M. M. L.; SÁ JUNIOR, J. J. M. Alternative signatures based on randomized neural network for texture classification. In: *2019 XV Workshop de Visão Computacional (WVC)*. [s.n.], 2019. p. 61–65. Disponível em: <https://doi.org/10.1109/WVC.2019.8876952>. Citado na página 17.

KANNALA, J.; RAHTU, E. BSIF: Binarized statistical image features. In: *ICPR*. IEEE Computer Society, 2012. p. 1363–1366. Disponível em: <https://ieeexplore.ieee.org/document/6460393>. Citado na página 67.

LAZEBNIK, S.; SCHMID, C.; PONCE, J. A sparse texture representation using local affine regions. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, v. 27, n. 8, p. 1265–1278, 2005. Disponível em: <https://doi.org/10.1109/TPAMI.2005.151>. Citado 4 vezes nas páginas 14, 47, 48 e 67.

LECUN, Y.; CORTES, C.; BURGES, C. J. C. The MNIST database of handwritten digits. In: . [s.n.], 1998. Disponível em: <http://yann.lecun.com/exdb/mnist/>. Citado na página 44.

LEUNG, T.; MALIK, J. Representing and recognizing the visual appearance of materials using three-dimensional textons. *International Journal of Computer Vision*, v. 43, n. 1, p. 29–44, Jun 2001. ISSN 1573-1405. Disponível em: <https://doi.org/10.1023/A:1011126920638>. Citado na página 28.

LIU, C.; YUEN, J.; TORRALBA, A. SIFT flow: Dense correspondence across scenes and its applications. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, v. 33, n. 5, p. 978–994, May 2011. ISSN 1939-3539. Disponível em: <https://doi.org/10.1109/TPAMI.2010.147>. Citado na página 28.

LIU, L.; FIEGUTH, P.; KUANG, G. Compressed sensing for robust texture classification. In: KIMMEL, R.; KLETTE, R.; SUGIMOTO, A. (Ed.). *Computer Vision – ACCV 2010*. Berlin, Heidelberg: Springer Berlin Heidelberg, 2011. p. 383–396. Disponível em: <https://doi.org/10.1007/978-3-642-19315-6_30>. Citado na página 67.

LOUART, C.; COUILLET, R. Harnessing neural networks: a random matrix approach. In: IEEE; Inst Elect & Elect Engineers, Signal Proc Soc. *2017 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*. 345 E 47TH ST, NEW YORK, NY 10017 USA: IEEE, 2017. (International Conference on Acoustics Speech and Signal Processing ICASSP), p. 2282–2286. ISBN 978-1-5090-4117-6. ISSN 1520-6149. IEEE International Conference on Acoustics, Speech, and Signal Processing (ICASSP), New Orleans, LA, MAR 05-09, 2017. Citado na página 43.

LOUART, C.; LIAO, Z.; COUILLET, R. A random matrix approach to neural networks. *Ann. Appl. Probab.*, The Institute of Mathematical Statistics, v. 28, n. 2,

p. 1190–1248, 04 2018. Disponível em: <https://doi.org/10.1214/17-AAP1328>. Citado 2 vezes nas páginas 41 e 43.

LU, S. et al. A pathological brain detection system based on kernel based elm. *Multimedia Tools and Applications*, v. 77, n. 3, p. 3715–3728, Feb 2018. ISSN 1573-7721. Disponível em: <https://doi.org/10.1007/s11042-016-3559-z>. Citado na página 19.

MANJUNATH, B.; MA, W. Texture features for browsing and retrieval of image data. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, IEEE Computer Society, v. 18, n. 8, p. 837–842, 1996. ISSN 0162-8828. Citado na página 89.

MELEKOODAPPATTU, J. G.; SUBBIAN, P. S. A hybridized elm for automatic micro calcification detection in mammogram images based on multi-scale features. *Journal of Medical Systems*, v. 43, n. 7, p. 183, May 2019. ISSN 1573-689X. Disponível em: <https://doi.org/10.1007/s10916-019-1316-3>. Citado na página 19.

MOORE, E. H. On the reciprocal of the general algebraic matrix. *Bull. Am. Math. Soc.*, v. 26, p. 394–395, 1920. Disponível em: <https://ci.nii.ac.jp/naid/10025388756/en/>. Citado na página 22.

NAYAK, D. R. et al. Deep extreme learning machine with leaky rectified linear unit for multiclass classification of pathological brain images. *Multimedia Tools and Applications*, SPRINGER, Van Godewijckstraat 30, 3311 GZ Dordrecht, Netherlands, 79, n. 21-22, p. 15381–15396, JUN 2020. ISSN 1380-7501. Disponível em: <https://doi.org/10.1007/s11042-019-7233-0>. Citado na página 19.

OJALA, T.; PIETIKÄINEN, M.; MÄENPÄÄ, T. Multiresolution gray-scale and rotation invariant texture classification with local binary patterns. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, v. 24, n. 7, p. 971–987, 2002. Disponível em: <https://doi.org/10.1109/TPAMI.2002.1017623>. Citado 5 vezes nas páginas 11, 26, 28, 29 e 67.

Ojala, T.; Pietikainen, M.; Maenpaa, T. Multiresolution gray-scale and rotation invariant texture classification with local binary patterns. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, v. 24, n. 7, p. 971–987, July 2002. ISSN 1939-3539. Citado 2 vezes nas páginas 31 e 89.

Pao, Y. .; Takefuji, Y. Functional-link net computing: theory, system architecture, and functionalities. *Computer*, v. 25, n. 5, p. 76–79, 1992. Citado na página 20.

PAO, Y.-H.; PARK, G.-H.; SOBAJIC, D. J. Learning and generalization characteristics of the random vector functional-link net. *Neurocomputing*, v. 6, n. 2, p. 163 – 180, 1994. ISSN 0925-2312. Backpropagation, Part IV. Disponível em: <http://www.sciencedirect.com/science/article/pii/0925231294900531>. Citado na página 20.

PEDRINI, H.; SCHWARTZ, W. R. *Análise de imagens digitais: príncipios, algoritmos e aplicações*. [S.l.]: São Paulo, 2008. Citado na página 28.

PENROSE, R. A generalized inverse for matrices. *Mathematical Proceedings of the Cambridge Philosophical Society*, Cambridge University Press, v. 51, n. 3, p. 406–413, 1955. Citado na página 22.

QUAN, Y. et al. Lacunarity analysis on image patterns for texture classification. In: *2014 IEEE Conference on Computer Vision and Pattern Recognition*. [s.n.], 2014. p. 160–167. Disponível em: <https://doi.org/10.1109/CVPR.2014.28>. Citado na página 67.

RIBAS, L. C. et al. Fusion of complex networks and randomized neural networks for texture analysis. *Pattern Recognition*, v. 103, p. 107189, 2020. ISSN 0031-3203. Disponível em: <http://www.sciencedirect.com/science/article/pii/S0031320319304893>. Citado 2 vezes nas páginas 17 e 38.

SÁ JUNIOR, J. J. M.; BACKES, A. R.; BRUNO, O. M. Pap-smear image classification using randomized neural network based signature. In: MENDOZA, M.; VELASTÍN, S. (Ed.). *Progress in Pattern Recognition, Image Analysis, Computer Vision, and Applications*. Cham: Springer International Publishing, 2018. p. 677–684. ISBN 978-3-319-75193-1. Disponível em: <https://doi.org/10.1007/978-3-319-75193-1_81>. Citado na página 18.

_____. Randomized neural network based signature for classification of titanium alloy microstructures. In: MENDOZA, M.; VELASTÍN, S. (Ed.). *Progress in Pattern Recognition, Image Analysis, Computer Vision, and Applications*. Cham: Springer International Publishing, 2018. p. 669–676. ISBN 978-3-319-75193-1. Disponível em: <https://doi.org/10.1007/978-3-319-75193-1_80>. Citado na página 18.

Schmidt, W. F.; Kraaijveld, M. A.; Duin, R. P. W. Feedforward neural networks with random weights. In: *Proceedings., 11th IAPR International Conference on Pattern Recognition. Vol.II. Conference B: Pattern Recognition Methodology and Systems*. [S.l.: s.n.], 1992. p. 1–4. Citado na página 20.

SENGUPTA, S. et al. A review of deep learning with special emphasis on architectures, applications and recent trends. *Knowledge-Based Systems*, Elsevier, Radarweg 29, 1043 NX Amsterdam, Netherlands, 194, APR 22 2020. ISSN 0950-7051. Disponível em: <https://doi.org/10.1016/j.knosys.2020.105596>. Citado na página 10.

SUSSNER, P.; CAMPIOTTI, I. Extreme learning machine for a new hybrid morphological/linear perceptron. *Neural Networks*, v. 123, p. 288 – 298, 2020. ISSN 0893-6080. Disponível em: <http://www.sciencedirect.com/science/article/pii/S0893608019303958>. Citado 6 vezes nas páginas 12, 14, 18, 35, 36 e 37.

SZELISKI, R. *Computer Vision: Algorithms and Applications*. [S.l.]: Berlin, 2010. Citado na página 10.

SÁ JUNIOR, J. J. M.; BACKES, A. R. ELM based signature for texture classification. *Pattern Recognition*, v. 51, p. 395 – 401, 2016. ISSN 0031-3203. Disponível em: <http://www.sciencedirect.com/science/article/pii/S0031320315003453>. Citado 12 vezes nas páginas 11, 12, 13, 16, 17, 18, 20, 22, 26, 27, 33 e 93.

SÁ JUNIOR, J. J. M.; BACKES, A. R.; BRUNO, O. M. Randomized neural network based descriptors for shape classification. *Neurocomputing*, v. 312, p. 201 – 209, 2018. ISSN 0925-2312. Disponível em: <http://www.sciencedirect.com/science/article/pii/S0925231218306842>. Citado na página 17.

SÁ JUNIOR, J. J. M.; RIBAS, L. C.; BRUNO, O. M. Randomized neural network based signature for dynamic texture classification. *Expert Systems with Applications*, v. 135, p. 194 – 200, 2019. ISSN 0957-4174. Disponível em: <http://www.sciencedirect.com/science/article/pii/S0957417419303914>. Citado 2 vezes nas páginas 17 e 44.

TIKHONOV, A. N. On the solution of ill-posed problems and the method of regularization. *Dokl. Akad. Nauk SSSR*, v. 151, n. 3, p. 501 – 504, 1963. Disponível em: <http://www.ams.org/mathscinet-getitem?mr=0162377>. Citado na página 38.

TURKOGLU, M.; HANBAY, D. Leaf-based plant species recognition based on improved local binary pattern and extreme learning machine. *Physica A: Statistical Mechanics and its Applications*, v. 527, p. 121297, 2019. ISSN 0378-4371. Disponível em: <http://www.sciencedirect.com/science/article/pii/S0378437119307551>. Citado na página 18.

VAPNIK, V. N. *The nature of statistical learning theory*. [S.l.]: New Youk, 2000. Citado na página 11.

VARMA, M.; ZISSERMAN, A. A statistical approach to texture classification from single images. *International Journal of Computer Vision*, v. 62, n. 1, p. 61–81, 2005. Disponível em: <https://doi.org/10.1023/B:VISI.0000046589.39864.ee>. Citado na página 67.

______. A statistical approach to material classification using image patch exemplars. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, v. 31, n. 11, p. 2032–2047, 2009. Disponível em: <https://doi.org/10.1109/TPAMI.2008.182>. Citado na página 67.

WAJID, S. K. et al. Lung cancer detection using local energy-based shape histogram (lesh) feature extraction and cognitive machine learning techniques. In: *2016 IEEE 15th International Conference on Cognitive Informatics Cognitive Computing (ICCI*CC)*. [s.n.], 2016. p. 359–366. ISSN null. Disponível em: <https://doi.org/10.1109/ICCI-CC.2016.7862060>. Citado na página 18.

WANG, Z. et al. Improved lung nodule diagnosis accuracy using lung ct images with uncertain class. *Computer Methods and Programs in Biomedicine*, v. 162, p. 197 – 209, 2018. ISSN 0169-2607. Disponível em:

<http://www.sciencedirect.com/science/article/pii/S0169260717311495>. Citado 2 vezes nas páginas 19 e 70.

WATKINS, D. S. *Fundamentals of Matrix Computations*. 2. ed. USA: John Wiley & Sons, Inc., 2002. ISBN 0471213942. Citado 2 vezes nas páginas 23 e 24.

XU, Y. et al. Combining powerful local and global statistics for texture description. In: IEEE. *IEEE Conference on Computer Vision and Pattern Recognition, 2009. CVPR 2009*. IEEE, 2009. p. 573 – 580. Disponível em: <https://doi.org/10.1109/CVPR.2009.5206741>. Citado na página 67.

XU, Y.; JI, H.; FERMÜLLER, C. Viewpoint invariant texture description using fractal analysis. *International Journal of Computer Vision*, v. 83, n. 1, p. 85–100, 2009. Disponível em: <https://doi.org/10.1007/s11263-009-0220-6>. Citado 4 vezes nas páginas 14, 44, 47 e 67.

ZHOU, C.; LI, F.; CAO, W. Architecture design and implementation of image based autonomous car: THUNDER-1. *Multimedia Tools and Applications*, Springer, Van Godewijckstraat 30, 3311 GZ Dordrecht, Netherlands, 78, n. 20, p. 28557–28573, OCT 2019. ISSN 1380-7501. Disponível em: <https://doi.org/10.1007/s11042-018-5816-9>. Citado na página 10.