

Instituto Federal de Educação, Ciência e Tecnologia do Ceará

Pró-Reitoria de Pesquisa, Pós-Graduação e Inovação

Programa de Pós-Graduação em Ciência da Computação

Luan Sousa Cordeiro

**Identificação automática de comprimidos utilizando aprendizado de máquina e
visão computacional**

Fortaleza – CE

2025

Luan Sousa Cordeiro

**Identificação automática de comprimidos utilizando aprendizado de máquina e
visão computacional**

Dissertação submetida ao Programa de Pós-Graduação em Ciência da Computação do Instituto Federal de Educação, Ciência e Tecnologia do Ceará, como requisito parcial para obtenção do grau de Mestre em Ciência da Computação.

Linha de pesquisa: Inteligência Artificial

Orientador: Prof. Dr. Francisco Nivando Bezerra
Coorientador: Prof. Dr. Saulo Anderson de Freitas
Oliveira

Instituto Federal de Educação, Ciência e Tecnologia do Ceará

Pró-Reitoria de Pesquisa, Pós-Graduação e Inovação

Programa de Pós-Graduação em Ciência da Computação

Fortaleza – CE

2025

Dados Internacionais de Catalogação na Publicação
Instituto Federal do Ceará - IFCE
Sistema de Bibliotecas - SIBI
Ficha catalográfica elaborada pelo SIBI/IFCE, com os dados fornecidos pelo(a) autor(a)

C794i Cordeiro, Luan Sousa.
 Identificação automática de comprimidos utilizando aprendizado de máquina e visão computacional /
 Luan Sousa Cordeiro. - 2025.
 97 f. : il. color.

 Dissertação (Mestrado) - Instituto Federal do Ceará, Mestrado em Ciência da Computação, Campus
Fortaleza, 2025.

 Orientação: Prof. Dr. Francisco Nivando Bezerra.

 Coorientação: Prof. Dr. Saulo Anderson de Freitas Oliveira.

 1. identificação de comprimidos. 2. aprendizagem de máquina. 3. visão computacional. 4. aprendizado
com poucos exemplos. 5. aprendizado autossupervisionado. I. Título.

CDD 004



Instituto Federal de Educação, Ciência e Tecnologia do Ceará
Pró-Reitoria de Pesquisa, Pós-Graduação e Inovação
Programa de Pós-Graduação em Ciência da Computação

Luan Sousa Cordeiro

Esta dissertação foi julgada adequada para a obtenção do título de Mestre em Ciência da Computação, sendo aprovada pela Coordenação do Programa de Pós-Graduação em Ciência da Computação do Instituto Federal de Educação, Ciência e Tecnologia do Ceará e pela banca examinadora:

Orientador: Prof. Dr. Francisco Nivando Bezerra

Instituto Federal do Ceará – IFCE

Coorientador: Prof. Dr. Saulo Anderson de Freitas Oliveira

Instituto Federal do Ceará – IFCE

Prof. Dr. José Daniel de Alencar Santos

Instituto Federal do Ceará – IFCE

Prof. Dr. Iális Cavalcante de Paula Júnior

Universidade Federal do Ceará – UFC

Prof. Dr. Aldísio Gonçalves Medeiros

Universidade Federal do Ceará – UFC

Fortaleza – CE

2025

Dedico este trabalho à minha esposa, Hannah.

Agradecimentos

Realizar uma pesquisa acadêmica de mestrado em dois anos parecia, à primeira vista, uma tarefa viável. Contudo, a cada dia que passava eu compreendia melhor que essa não era apenas uma tarefa difícil – era um desafio que, sem o devido suporte e direcionamento, se impossibilitaria de realizar. Por isso, reservo este espaço para agradecer a todos que contribuíram, de alguma forma, para eu alcançar esta conquista.

Agradeço, antes de tudo, a Deus por ser minha força e serenidade nos momentos em que o cansaço ameaçou silenciar o meu entusiasmo. Sem dúvida alguma, a Ele devemos toda honra e toda glória.

Agradeço à minha esposa, Hannah, por me apoiar diariamente, por compreender os momentos em que estive ausente, por ser a pessoa com quem eu tenho o privilégio de dividir os meus dias e por ser minha inspiração diária. Sem você, eu não teria chegado onde cheguei. Essa conquista também é sua.

Ao meu orientador, prof. Nivando, agradeço por aceitar me orientar, por sempre estar disponível para tirar as minhas dúvidas – mesmo em períodos de férias – pela confiança que depositou em mim, e por me ensinar que a pesquisa científica é, acima de tudo, um exercício de humildade intelectual. A sua dedicação e olhar atento foram determinantes para que eu chegasse a este momento.

Ao meu coorientador, prof. Saulo, que inicialmente participou como membro da banca de qualificação e aceitou contribuir ativamente para o desenvolvimento desta pesquisa, agradeço pelas inúmeras discussões que enriqueceram profundamente este trabalho.

Aos membros da banca avaliadora, agradeço por dedicarem o seu tempo para lapidar esta dissertação, com comentários completos, complementares e precisos. As suas observações e questionamentos contribuíram significativamente para o amadurecimento deste trabalho e ampliaram a minha visão sobre o tema de pesquisa.

À minha família, base e abrigo em todas as fases da vida. Aos meus pais, pelo amor incondicional que me deram desde sempre. Sem vocês, eu não seria quem sou. Aos meus irmãos, que, junto dos nossos pais, são uma das razões pelas quais eu tanto prezo a vida.

Agradeço também ao meu líder no trabalho, Kildere, por apoiar o meu desejo de cursar o mestrado. Ao longo dos quase cinco anos em que estamos a trabalhar juntos, adquiri uma grande admiração profissional e pessoal por você.

E, por fim, agradeço a mim mesmo, por seguir mesmo quando o caminho parecia árduo demais. Esta dissertação é também uma prova de que persistir vale a pena.

*“A adversidade desperta em nós capacidades que, em circunstâncias
favoráveis, teriam ficado adormecidas.”*
(Horácio)

Resumo

A identificação correta de comprimidos é fundamental para prevenir erros de medicação, que representam um grave risco à segurança do paciente. Embora a agência reguladora FDA (*Food and Drug Administration*) exija uma aparência visual única para cada comprimido, a complexidade e a vasta quantidade de fármacos no mercado tornam a identificação manual uma tarefa desafiadora, evidenciando a necessidade de soluções automatizadas. Diante desse cenário, esta dissertação desenvolve e avalia soluções de visão computacional para a identificação automática de comprimidos, explorando desde abordagens tradicionais até técnicas avançadas de aprendizado profundo. A primeira metodologia foca na identificação de formas, empregando processamento de imagem para segmentar os comprimidos e extrair um conjunto de atributos geométricos de baixa complexidade computacional, como circularidade, solidez e alongamento. Esses atributos são utilizados para treinar múltiplos classificadores de aprendizado de máquina, incluindo *Support Vector Machine* (SVM), *Linear Discriminant Analysis* (LDA) e *k-Nearest Neighbours* (*k*-NN), com avaliação realizada sobre as imagens de referência do conjunto de dados NLM PIR. Os resultados demonstraram alta eficácia, com o classificador SVM de *kernel* linear alcançando a maior acurácia média de 95,71%. A segunda contribuição deste trabalho é a metodologia SSL4PILL, um modelo inovador voltado à identificação de comprimidos em cenários com poucos dados rotulados. Estruturada em dois estágios, a abordagem inicialmente utiliza aprendizado autossupervisionado para treinar um codificador baseado no modelo *Augmented Multiscale Deep InfoMax* (AMDIM), empregando um conjunto de imagens genéricas não rotuladas para aprender representações visuais robustas e generalizáveis. Em seguida, o modelo é refinado para a tarefa de identificação de comprimidos com o modelo *Prototypical Networks* (ProtoNet), que aprende a classificar novas classes de comprimidos a partir de um número mínimo de exemplos rotulados. Avaliamos o SSL4PILL nas imagens de consulta dos conjuntos de dados NLM PIR e CURE, o qual obteve desempenhos relevantes, com acurácia de 83,72% no NLM PIR, e acurácia de 60,78% no CURE em uma configuração de *1-shot 5-way*. Um estudo de ablação confirmou o papel crucial do pré-treinamento autossupervisionado, cuja remoção resultou em queda significativa no desempenho. Em conclusão, esta pesquisa demonstrou a aplicabilidade de diferentes paradigmas da visão computacional para a automação da identificação de comprimidos. As metodologias propostas não apenas alcançaram alto desempenho em suas respectivas tarefas, mas também forneceram entendimentos valiosos sobre a eficácia de abordagens de baixa e alta complexidade. Adicionalmente, elas propuseram uma solução promissora para a limitação de dados rotulados no domínio biomédico.

Palavras-chave: identificação de comprimidos. visão computacional. aprendizado de máquina. aprendizado com poucos exemplos. aprendizado autossupervisionado.

Abstract

The correct identification of pills is essential to prevent medication errors, which pose a serious risk to patient safety. Although the FDA (Food and Drug Administration) requires a unique visual appearance for each pill, the complexity and vast number of drugs on the market make manual identification a challenging task, highlighting the need for automated solutions. Given this scenario, this dissertation develops and evaluates computer vision solutions for automatic pill identification, exploring everything from classical approaches to advanced deep learning techniques. The first methodology focuses on identifying shapes, using image processing to segment the pills and extract a set of geometric attributes of low computational complexity, such as roundness, solidity, and elongation. These attributes are used to train multiple machine learning classifiers, including Support Vector Machine (SVM), Linear Discriminant Analysis (LDA), and k -Nearest Neighbors (k -NN), with evaluation performed on reference images from the NLM PIR dataset. The results demonstrated high effectiveness, with the linear kernel SVM classifier achieving the highest average accuracy of 95.71%. The second contribution of this work was the SSL4PILL methodology, an innovative model aimed at identifying pills in scenarios with limited labeled data. Structured in two stages, the approach initially uses self-supervised learning to train an encoder based on the Augmented Multiscale Deep InfoMax (AMDIM) model, employing a set of unlabeled generic images to learn robust and generalizable visual representations. Then, the model is refined for the pill identification task using the Prototypical Networks (ProtoNet) model, which learns to classify new pill classes from a minimum number of labeled examples. We evaluated SSL4PILL on consumer images from the NLM PIR and CURE datasets, achieving relevant performances, with 83.72% accuracy on NLM PIR, and 60.78% accuracy on CURE in a 1-shot 5-way configuration. An ablation study confirmed the crucial role of self-supervised pre-training, the removal of which resulted in a significant drop in performance. In conclusion, this research demonstrated the applicability of different computer vision paradigms for automating pill identification. The proposed methodologies not only performed well on their respective tasks, but also provided valuable insights into the effectiveness of low- and high-complexity approaches. Additionally, they propose a promising solution to the limitation of labeled data in the biomedical domain.

Keywords: pill identification. computer vision. machine learning. few-shot learning. self-supervised learning.

Lista de ilustrações

Figura 1 – Relação entre erros de medicação, eventos adversos a medicamentos e potenciais eventos adversos a medicamentos.	19
Figura 2 – Linha do tempo da literatura acadêmica sobre identificação de comprimidos. Legenda de cores: ■ visão computacional tradicional, ■ transição para aprendizado de máquina, ■ CNN clássicas, ■ aprendizado com poucos exemplos, ■ detecção, ■ métodos híbridos, ■ dados multimodais.	33
Figura 3 – Exemplos de imagens de referência e de consulta para um mesmo comprimido do conjunto de dados NLM PIR. Da esquerda para a direita: imagem de referência (verso do comprimido), imagem de consulta 1 e imagem de consulta 2. Os nomes originais das amostras estão listados no Apêndice D.	35
Figura 4 – Mosaico com 12 imagens de referência do conjunto de dados NLM PIR. Os nomes originais das amostras estão listados no Apêndice D.	36
Figura 5 – Exemplos de imagens de referência e de consulta para um mesmo comprimido do conjunto de dados CURE. Da esquerda para a direita: imagem de referência (verso do comprimido), imagem de consulta 1 e imagem de consulta 2. Os nomes originais das amostras estão listados no Apêndice D.	37
Figura 6 – Mosaico com 12 imagens de consulta do conjunto de dados CURE. Os nomes originais das amostras estão listados no Apêndice D.	37
Figura 7 – Exemplos de amostras difíceis do conjunto de dados CURE. A primeira linha ilustra positivos difíceis. A segunda linha apresenta negativos difíceis. Os nomes originais das amostras estão listados no Apêndice D.	38
Figura 8 – Ilustração das subetapas do processo de segmentação em imagens do conjunto de dados NLM PIR. Na primeira linha, estão as imagens de entrada. Na segunda linha, estão as imagens em níveis de cinza. Na terceira linha, estão as imagens segmentadas. Os nomes originais das amostras estão listados no Apêndice D.	39
Figura 9 – Fluxo de processamento para identificação de formas de comprimidos.	42
Figura 10 – Distribuição das formas dos comprimidos do conjunto de dados NLM PIR. Para uma melhor visualização, o eixo horizontal está em escala logarítmica.	43
Figura 11 – Matrizes de confusão em que os classificadores LDA e QDA obtiveram maior acurácia no conjunto de teste.	52
Figura 12 – Matrizes de confusão em que os classificadores NB e k -NN obtiveram maior acurácia no conjunto de teste.	52

Figura 13 – Matrizes de confusão em que os classificadores MLP e ELM obtiveram maior acurácia no conjunto de teste.	53
Figura 14 – Matrizes de confusão em que os classificadores SVM/LIN, SVM/POL e SVM/RBF obtiveram maior acurácia no conjunto de teste.	54
Figura 15 – Estrutura da metodologia SSL4PILL para identificação de comprimidos com poucos exemplos e autossupervisão.	60
Figura 16 – (a) Processo de aumento de dados que gera duas visualizações aumentadas $\mathbf{x}_a$ e $\mathbf{x}_b$ a partir de uma imagem original $\mathbf{x}$. (b) Arquitetura do AMDIM, com função de custo mustiescala em visões geradas por aumento de dados.	62
Figura 17 – Etapa de refinamento da metodologia SSL4PILL com ProtoNet.	64
Figura 18 – Ilustração conceitual do classificador NB. Cada atributo é modelado por uma distribuição condicional independente, e as probabilidades são combinadas para inferir a classe mais provável.	84
Figura 19 – Arquitetura típica de um classificador MLP.	85
Figura 20 – Esquema de busca em grade com validação cruzada. Cada ponto da grade representa uma combinação de hiperparâmetros avaliada sistematicamente.	90
Figura 21 – Uso da técnica <i>holdout</i> para estimação do risco estrutural de um modelo.	91
Figura 22 – Uso da técnica <i>k-fold</i> para estimação do risco estrutural de um modelo.	91

Lista de tabelas

Tabela 1	– Características dos conjuntos de dados NLM PIR e CURE.	38
Tabela 2	– Atributos de forma.	40
Tabela 3	– Valores dos atributos de forma para as imagens binárias na Figura 8. .	41
Tabela 4	– Hiperparâmetros dos classificadores k -NN, MLP, ELM, SVM/LIN, SVM/POL e SVM/RBF.	47
Tabela 5	– Resultados dos classificadores no problema de identificação de formas, para as métricas: acurácia, MCC, precisão e revocação. Os valores de cada métrica são descritos em termos da média e desvio padrão, após 30 realizações de treinamento e teste. Destacamos os melhores desempenhos com fonte em negrito e cor de fundo laranja.	48
Tabela 6	– Resultados, por classe, das métricas precisão e revocação. Os valores de cada métrica são descritos em termos da média e desvio padrão, após 30 realizações de treinamento e teste. Destacamos os melhores desempenhos com fonte em negrito e cor de fundo laranja.	50
Tabela 7	– Resultados do teste estatístico de Wilcoxon, com base nos valores de acurácia (%), para aferição de equivalência ou não equivalência entre o método que propomos e os extratores: Momentos Estatísticos (Mom. Es.), Momentos Centrais (Mom. Ce.), Momentos de Hu (Mom. Hu), LBP e GLCM. Os símbolos ✓ e ✗ indicam equivalência ou não equivalência, respectivamente.	56
Tabela 8	– Resultados do teste estatístico de Wilcoxon, com base nos valores do MCC, para aferição de equivalência ou não equivalência entre o método que propomos e os extratores: Momentos Estatísticos (Mom. Es.), Momentos Centrais (Mom. Ce.), Momentos de Hu (Mom. Hu), LBP e GLCM. Os símbolos ✓ e ✗ indicam equivalência ou não equivalência, respectivamente.	57
Tabela 9	– Acurácia média (%) na configuração 1-shot 5-way nos conjuntos de dados NLM PIR e CURE. Ling <i>et al.</i> (2020) e Chee <i>et al.</i> (2024a) não forneceram intervalos de confiança. Destacamos as maiores acurácias em negrito e cor de fundo laranja.	67
Tabela 10	– Nomes dos arquivos utilizados na Figura 3. A posição indica a ordem das imagens, da esquerda para a direita.	94
Tabela 11	– Nomes dos arquivos utilizados na Figura 4. A posição indica a ordem das imagens, da esquerda para a direita e de cima para baixo.	95
Tabela 12	– Nomes dos arquivos utilizados na Figura 5. A posição indica a ordem das imagens, da esquerda para a direita.	95

Tabela 13 – Nomes dos arquivos utilizados na Figura 6. A posição indica a ordem das imagens, da esquerda para a direita e de cima para baixo.	96
Tabela 14 – Nomes dos arquivos utilizados na Figura 7. A posição indica a ordem das imagens, da esquerda para a direita e de cima para baixo.	96
Tabela 15 – Nomes dos arquivos utilizados na Figura 8. A posição indica a ordem das imagens, da esquerda para a direita.	97

Lista de abreviaturas e siglas

ADE	<i>Adverse Drug Event</i>
AG	Alongamento
CL	Circularidade
CNN	<i>Convolutional Neural Network</i>
CP	Compacidade
CTM	<i>Category Traversal Module</i>
C3PI	<i>Computational Photography Project for Pill Identification</i>
ELM	<i>Extreme Learning Machine</i>
EUA	Estados Unidos da América
FDA	<i>Food and Drug Administration</i>
FN	Falso Negativo
FP	Falso Positivo
FPS	<i>Frames Per Second</i>
GAP	<i>Global Average Precision</i>
GLCM	<i>Gray Level Co-occurrence Matrix</i>
IoT	<i>Internet of Things</i>
k -NN	<i>K-Nearest Neighbors</i>
LBP	<i>Local Binary Patterns</i>
LDA	<i>Linear Discriminant Analysis</i>
MAP	<i>Mean Average Precision</i>
MCC	<i>Matthews Correlation Coefficient</i>
MLP	<i>Multilayer Perceptron</i>
MPI	<i>Multi-Pill Identifier</i>
MS	<i>Multi Stream</i>

MTL	<i>Meta-Transfer Learning</i>
NB	<i>Naïve Bayes</i>
NDC	<i>National Drug Code</i>
NLM	<i>National Library of Medicine</i>
OCR	<i>Optical Character Recognition</i>
PA	Razão Perímetro-Área
PIR	<i>Pill Image Recognition</i>
QDA	<i>Quadratic Discriminant Analysis</i>
RD	Redondeza
RGB	<i>Red, Green, Blue</i>
SGD	<i>Stochastic Gradient Descent</i>
SIFT	<i>Scale-Invariant Feature Transform</i>
SL	Solidez
SMO	<i>Sequential Minimal Optimization</i>
SSD	<i>Single Shot MultiBox Detector</i>
SVM	<i>Support Vector Machine</i>
TSDS	<i>Two-Step Sampling Distance Sets</i>
VN	Verdadeiro Negativo
VP	Verdadeiro Positivo
YOLO	<i>You Only Look Once</i>

Sumário

1	INTRODUÇÃO	18
1.1	Motivações da dissertação	19
1.2	Objetivo geral	21
1.3	Objetivos específicos	21
1.4	Estrutura da dissertação	21
1.5	Produção científica	22
2	REVISÃO DE LITERATURA	23
2.1	Visão computacional tradicional	23
2.2	Métodos estatísticos e aprendizado de máquina	25
2.3	Redes neurais convolucionais	26
2.4	Aprendizado com poucos exemplos	28
2.5	Arquiteturas com múltiplos fluxos e comitês de classificadores	29
2.6	Detecção e reconhecimento simultâneos	30
2.7	Contribuições teóricas e interdomínios	31
2.8	Considerações finais	32
3	CONJUNTOS DE DADOS	34
3.1	NLM PIR	34
3.2	CURE	35
3.3	Uma estratégia preliminar para identificação de comprimidos	39
3.3.1	Segmentação	39
3.3.2	Extração de atributos	40
4	IDENTIFICAÇÃO DE FORMAS DE COMPRIMIDOS COM TÉCNICAS DE PROCESSAMENTO DE IMAGENS	42
4.1	Análise das formas dos comprimidos	43
4.2	Experimentos e discussões	44
4.2.1	Projeto dos classificadores	45
4.2.2	Análise geral dos resultados	47
4.2.3	Análise individual dos classificadores	51
4.2.4	Comparação com outros extratores de atributos	55
4.3	Considerações finais	58
5	IDENTIFICAÇÃO DE COMPRIMIDOS COM POUCOS EXEM- PLOS E AUTOSSUPERVISÃO	59

5.1	SSL4PILL	60
5.1.1	Estágio 1: pré-treinamento com autossupervisão	60
5.1.2	Estágio 2: refinamento com aprendizado de poucos exemplos	63
5.2	Configuração experimental	64
5.2.1	Conjuntos de dados	64
5.2.2	Pré-treinamento com AMDIM	65
5.2.3	Ajuste fino com ProtoNet	65
5.2.4	Ambiente computacional	66
5.3	Avaliação e resultados	66
5.3.1	Protocolos de avaliação	66
5.3.2	Resultados quantitativos	66
5.3.3	Estudo de ablação: impacto do pré-treinamento	67
5.3.4	Discussão dos resultados	68
5.3.5	Limitações e generalização	68
6	CONCLUSÕES E TRABALHOS FUTUROS	69
	REFERÊNCIAS	71
	APÊNDICES	79
	APÊNDICE A – CLASSIFICADORES PARA IDENTIFICAÇÃO DE FORMAS DE COMPRIMIDOS	80
A.1	Classificadores baseados em distância	80
A.1.1	<i>k</i> -Nearest Neighbors (<i>k</i> -NN)	80
A.1.2	<i>Linear Discriminant Analysis</i> (LDA)	80
A.1.2.1	Interpretação geométrica do LDA	81
A.1.3	<i>Quadratic Discriminant Analysis</i> (QDA)	81
A.1.3.1	Interpretação geométrica do QDA	82
A.2	Classificadores baseados em probabilidade	82
A.2.1	<i>Naïve Bayes</i> (NB)	83
A.3	Classificadores baseados em funções	84
A.3.1	Redes neurais artificiais	84
A.3.1.1	<i>Multilayer Perceptron</i> (MLP)	85
A.3.1.2	<i>Extreme Learning Machine</i> (ELM)	86
A.3.2	<i>Support Vector Machine</i> (SVM)	86
A.3.2.1	Margens Rígidas	87
A.3.2.2	Margens Suaves	88

	APÊNDICE B – BUSCA EM GRADE E TÉCNICAS DE VALIDAÇÃO CRUZADA	89
B.1	Busca em grade	89
B.2	Validação cruzada	90
B.2.1	<i>Holdout</i>	90
B.2.2	<i>k-fold</i>	91
	APÊNDICE C – MÉTRICAS DE DESEMPENHO	92
	APÊNDICE D – IDENTIFICAÇÃO DAS IMAGENS UTILIZADAS NO CAPÍTULO 3	94

1 Introdução

A partir do ano 2000, o Instituto de Medicina (*Institute of Medicine*) dos Estados Unidos da América (EUA) publicou uma série de relatórios que tornou público vários problemas de segurança médica (BATES, 2007). Dessa série, o relatório intitulado “To Err Is Human” é talvez o mais importante, pois criou um marco transformacional para segurança de pacientes (BATES; SINGH, 2018). O relatório destaca várias questões importantes, entre elas: erros são comuns e custosos, problemas relacionados a sistemas causam erros, erros são preveníveis, e pode-se melhorar a segurança do paciente (BATES; SINGH, 2018 apud MEDICINE, 2000).

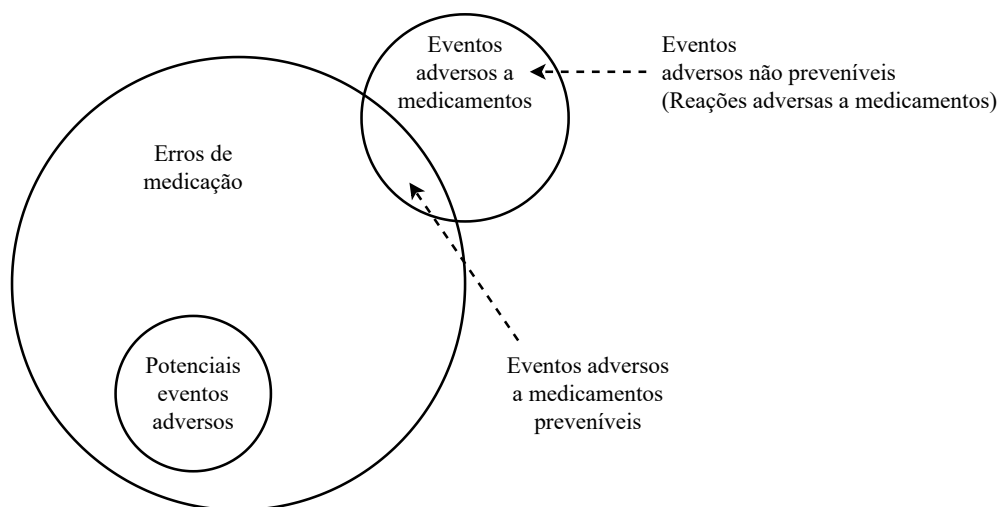
À época, o relatório estimou que erros médicos causavam de 44 mil a 98 mil mortes por ano nos EUA. Embora alguns estudos afirmem que esses números são exagerados (BRENNAN, 2000; MCDONALD; WEINER; HUI, 2000), outros reforçam e até aumentam o número de mortes por ano. É o caso do artigo publicado por Makary e Daniel (2016), em que os autores argumentam que os números atuais se aproximam de 400 mil mortes por ano, tornando o erro médico a terceira maior causa de morte nos EUA. Apesar das controvérsias, o relatório não só permaneceu como marco na área de segurança em medicina, como abriu espaço para novos estudos e melhorias no sistema de saúde.

O relatório aponta ainda que erros relacionados a medicamentos são o tipo mais prevenível de erro médico (MEDICINE, 2000). Erros de medicação podem ocorrer em qualquer estágio do processo de uso da medicação e podem ou não acarretar eventos adversos¹ (*adverse drug events*, ADE). Dependendo do cenário clínico, cerca de um terço a metade dos ADEs estão associados a erros de medicação. A relação entre ADEs, potenciais ADEs e erros de medicação é mostrada na Figura 1.

Erros na prescrição ou administração de um comprimido podem levar a eventos adversos a medicamentos. A identificação correta dos comprimidos de medicamentos tem o potencial de melhorar o tratamento, desde a experiência do paciente até os centros de controle de intoxicações. A Administração Federal de Alimentos e Medicamentos (*Food and Drug Administration*, FDA), por meio de seu código regulatório 21CFR206, impõe uma aparência única para cada prescrição em termos de tamanho, formato, cor e impressão. No entanto, a lei de propriedade intelectual dos EUA (chamada de “trade dress”) permite situações em que um único comprimido tenha aparências diferentes (GREENE; KESSELHEIM, 2011). Portanto, identificar comprimidos com base em sua aparência visual torna-se um desafio que requer algum grau de instrução em saúde.

¹ Um evento adverso é uma “lesão resultante do tratamento médico, e não da condição subjacente do paciente”. Em outras palavras, refere-se a qualquer dano indesejado provocado diretamente por um medicamento, procedimento ou terapia, e não pela doença que está sendo tratada (MEDICINE, 2007).

Figura 1 – Relação entre erros de medicação, eventos adversos a medicamentos e potenciais eventos adversos a medicamentos.



Fonte: Adaptação de [Gandhi, Seder e Bates \(2000\)](#).

1.1 Motivações da dissertação

Identificar corretamente comprimidos de uso oral (também chamados de pílulas), baseado em suas aparências visuais, é um passo essencial para assegurar a segurança do paciente e possibilitar um atendimento mais eficaz.

Do ponto de vista da segurança, uma variedade de erros pode ocorrer ao longo da cadeia farmacológica, incluindo falhas na prescrição, transcrição (interpretação da ordem de medicação), distribuição e administração ([BEMT; EGBERTS, 2007](#)). Tais erros podem ocasionar ADEs, ou seja, lesões resultantes do uso de medicamentos. A identificação e a confirmação de comprimidos prescritos podem mitigar erros de distribuição. Embora as taxas desses erros sejam baixas, drogarias continuam a incorporar tecnologias, como códigos de barras, para reduzi-las ainda mais ([CHEUNG; BOUVY; SMET, 2009](#)). A identificação visual é uma tecnologia adicional que aborda essa questão e que pode ser implantada também fora de drogarias, para uso anterior à administração do medicamento. Essa tecnologia é potencialmente aplicável tanto à área da saúde quanto ao público geral.

Sob a perspectiva do cuidado ao paciente, a identificação visual de comprimidos tem o potencial de melhorar o tratamento em diversas situações, especialmente quando há incerteza sobre o medicamento administrado ou ausência de informações precisas sobre a prescrição. Esse tipo de recurso é particularmente útil em contextos nos quais a assistência médica ocorre sob condições adversas, com restrições de tempo ou recursos, como em desastres naturais, emergências em locais públicos ou atendimentos domiciliares. Além disso, pode auxiliar profissionais de saúde na triagem inicial de pacientes, apoiar o trabalho

de centros de controle de intoxicações e favorecer a adesão a tratamentos de longo prazo.

Após desastres naturais ou causados pelo ser humano, é comum que comprimidos e documentação sejam separados. Uma revisão recente da literatura identificou a perda de medicações e a dificuldade em preencher prescrições como fenômenos recorrente em cenários de evacuação, nos quais os evacuados conseguem fornecer apenas informações parciais sobre suas medicações (OCHI *et al.*, 2014). Essa situação é especialmente preocupante para pessoas com condições crônicas, que podem sofrer efeitos adversos devido à interrupção em seus regimes de medicação. A importância de identificar medicamentos e preencher informações ausentes também foi evidenciada no fluxo operacional de um centro de evacuação bem-sucedido, implementado durante o furacão Katrina (IRVIN; ATAS, 2007), no qual o exame médico inicial incluía a identificação do medicamento por um farmacêutico.

Outro aspecto relevante do atendimento ao paciente é a atuação dos centros de controle de intoxicações. Nos EUA, esses centros oferecem consultoria especializada por telefone sobre possíveis exposições a substâncias tóxicas, incluindo medicamentos prescritos. O serviço é acessível tanto a profissionais de saúde quanto ao público em geral. A análise do volume de chamadas a esses centros demonstrou aumento significativo, principalmente em solicitações de identificação de comprimidos (SPILLER; GRIFFITH, 2009; KRENZELOK; MRVOS, 2009). Esse crescimento impôs um fardo considerável, pois ocupa o tempo de um número limitado de especialistas, desviando-os de atendimentos mais críticos. Uma solução eficaz foi descrita por Krenzelok e Mrvos (2011): o uso de um sistema interativo de resposta por voz para identificação automatizada de comprimidos reduziu em mais de 50% o volume de chamadas que exigiam intervenção humana, permitindo que os especialistas se dedicassem a casos mais urgentes.

Por fim, a identificação de comprimidos pode contribuir para aumentar a persistência do paciente, definida como “o tempo decorrido desde o início até a descontinuação da terapia” (CRAMER *et al.*, 2008), especialmente quando há alteração na aparência visual do medicamento prescrito. Essa questão é particularmente relevante para pacientes com condições crônicas, como hipertensão ou epilepsia, pois é provável que enfrentem essas mudanças diversas vezes ao longo do tratamento. Modificações de cor ou forma geralmente ocorrem em virtude da substituição entre versões de marca e genéricas ou entre genéricos distintos. Incentivos financeiros também podem provocar trocas frequentes entre genéricos equivalentes do mesmo medicamento.

Diversos estudos investigaram os efeitos dessas mudanças na persistência do paciente. Kesselheim *et al.* (2013, 2014) observaram redução na persistência quando os pacientes alternaram entre versões visualmente distintas de um mesmo medicamento. Por outro lado, Wijk *et al.* (2006) e Ude *et al.* (2011) não identificaram esse efeito, uma vez que as trocas avaliadas foram iniciadas por médicos e concentraram-se na substituição inicial de uma marca por um genérico. Finalmente, outras pesquisas apresentaram resultados mistos:

parte dos pacientes manteve ou aumentou a persistência, enquanto outros apresentaram redução (STEIN *et al.*, 2015; RATHE *et al.*, 2015). A persistência aumentada foi associada a menores custos e à familiaridade prévia com trocas de produtos; a redução, a pacientes que experimentaram sua primeira alteração. Assim, sistemas de identificação de comprimidos podem oferecer segurança adicional, assegurando aos pacientes que estão utilizando o medicamento correto e, potencialmente, promovendo maior adesão ao tratamento.

1.2 Objetivo geral

O objetivo desta dissertação é investigar estratégias de visão computacional para identificação automática de comprimidos, analisando técnicas de extração de atributos e métodos de aprendizado de máquina em contextos controlados e não controlados.

1.3 Objetivos específicos

De forma mais específica, esta dissertação tem os seguintes objetivos:

- a) analisar técnicas de extração de atributos a partir de características de forma em imagens de comprimidos obtidas em condições controladas;
- b) avaliar diferentes classificadores individuais na discriminação de formas de comprimidos, utilizando diversos extratores de atributos tradicionais;
- c) examinar técnicas de aprendizado profundo baseadas em autossupervisão e poucos exemplos para identificação de comprimidos, considerando imagens obtidas em condições não controladas;
- d) comparar o desempenho das metodologias propostas com soluções da literatura utilizando métricas de avaliação adequadas.

1.4 Estrutura da dissertação

Esta dissertação está estruturada em seis capítulos, além de apêndices e referências bibliográficas. A seguir, descrevemos o conteúdo de cada capítulo.

No Capítulo 2, apresentamos a revisão de literatura, abordando os principais conceitos e técnicas relacionados ao tema de pesquisa. Inicialmente, discutimos abordagens de visão computacional tradicionais e métodos estatísticos de aprendizado de máquina. Em seguida, apresentamos as redes neurais convolucionais e técnicas de aprendizado com poucos exemplos. Por fim, ainda exploramos arquiteturas com múltiplos fluxos e comitês de classificadores, detecção e reconhecimento simultâneos, contribuições teóricas e interdomínios relevantes para a identificação automática de comprimidos.

No Capítulo 3, descrevemos os conjuntos de dados utilizados na pesquisa, incluindo o NLM PIR e o CURE, detalhando suas características e aplicabilidade ao problema investigado. Adicionalmente, apresentamos uma estratégia preliminar para identificação de comprimidos, abordando técnicas de segmentação e extração de atributos.

No Capítulo 4, focamos na identificação de formas de comprimidos utilizando técnicas tradicionais de extração de atributos. Inicialmente, realizamos uma análise das formas dos comprimidos, seguida de experimentos detalhados que incluem o projeto dos classificadores, análise geral e individual dos resultados, e comparação com outros extratores de atributos disponíveis na literatura.

No Capítulo 5, abordamos a identificação de comprimidos utilizando aprendizado com poucos exemplos e autossupervisão. Propomos a metodologia SSL4PILL, que combina pré-treinamento com autossupervisão e ajuste fino com aprendizado de poucos exemplos. Em seguida, detalhamos a configuração experimental, protocolos de avaliação, resultados quantitativos, estudos de ablação e discussão das limitações e generalização da metodologia.

No Capítulo 6, apresentamos as conclusões da pesquisa, sintetizando os principais resultados obtidos e suas contribuições para a área de identificação automática de comprimidos. Por fim, sugerimos caminhos para trabalhos futuros que possam dar continuidade aos estudos realizados.

Os Apêndices fornecem informações complementares essenciais para a compreensão completa do trabalho. O Apêndice A descreve os classificadores que usamos para identificação de formas de comprimidos, organizados por categorias. O Apêndice B apresenta os métodos de busca em grade e validação cruzada que empregamos nos experimentos. O Apêndice C define as métricas de desempenho que utilizamos para avaliação das metodologias propostas. Finalmente, o Apêndice D lista os nomes dos arquivos das imagens de comprimidos presentes no Capítulo 3.

1.5 Produção científica

Essa pesquisa de mestrado gerou dois artigos de conferência, em momentos bem distintos. A seguir, estão as citações (oficial e preliminar) de ambos os trabalhos:

- 1 CORDEIRO, L. S. et al. Pill Image Classification using Machine Learning. In: IEEE 2019 8th Brazilian Conference on Intelligent Systems (BRACIS). [S.l.], 2019. p. 556–561. Disponível em: <<https://ieeexplore.ieee.org/abstract/document/8923984>>. Acesso em: 4 abr. 2023.
- 2 CORDEIRO, L. S, OLIVEIRA, S. A. F., BEZERRA, F. N. Self-supervised few-shot pill image recognition In: Intelligent Systems. BRACIS 2025. Lecture Notes in Computer Science. Springer, Cham.

2 Revisão de literatura

Atualmente, há uma variedade de soluções comerciais e serviços baseados na web para identificação de comprimidos. No entanto, essas opções ainda carecem de uma solução completa, prontamente disponível e fácil de usar, tanto para profissionais da saúde quanto para o público em geral, especialmente em contextos móveis ou de campo. A Biblioteca Nacional de Medicina (*National Library of Medicine*, NLM) dos EUA mantém o [DailyMed](#) (2025), voltado a profissionais de saúde, que exibe informações de rotulagem fornecidas pelo fabricante do comprimido e recuperadas pelo sistema com base no Código Nacional de Medicamentos (*National Drug Code*, NDC) do produto ou em palavras-chave.

Já a Pillbox ([MEDICINE](#), 2020), outra ferramenta anteriormente disponibilizada pela NLM, era mais apropriada para o público em geral, pois exibia informações com base em texto, como formato ou cor do comprimido. Embora a NLM tenha descontinuado a Pillbox, ela influenciou diversos outros sites que oferecem funcionalidades semelhantes, incluindo, entre outros, [WebMD](#) (2025), [RxList](#) (2025) e [Drugs](#) (2025). A principal desvantagem de todos esses sites é que a entrada é textual e cabe ao usuário selecionar as características corretas do comprimido, tarefa que permanece desafiadora (por exemplo, qual cor se deve selecionar de uma cápsula com duas cores?).

As limitações das soluções comerciais quanto à entrada visual e à usabilidade em cenários práticos evidenciam a necessidade de abordagens mais automatizadas e precisas. Nesse contexto, a literatura científica propôs diferentes estratégias computacionais para lidar com a variabilidade das condições de captura e características visuais dos comprimidos. Inicialmente, essas abordagens se basearam em técnicas tradicionais de visão computacional ([LEE et al.](#), 2012; [CHEN et al.](#), 2012; [YU et al.](#), 2015), sendo posteriormente substituídas ou complementadas por métodos baseados em aprendizado profundo ([ZENG; CAO; ZHANG](#), 2017; [DELGADO et al.](#), 2019; [LING et al.](#), 2020), com foco em maior robustez e escalabilidade. As seções seguintes apresentam essa trajetória, destacando os principais avanços e contribuições.

2.1 Visão computacional tradicional

Os primeiros trabalhos em reconhecimento de comprimidos estabeleceram fluxos de processamento (*pipelines*) de visão computacional tradicional, baseados em extração de contornos, descritores locais e correspondência de padrões. [Lee et al.](#) (2012) introduziram o PILL-ID, que utiliza descritores de bordas e correspondência de descritores SIFT (*Scale-Invariant Feature Transform*) para analisar e comparar impressões (*imprints*) gravadas na superfície dos comprimidos. Essa abordagem se destacou por sua robustez à variação

de escala e rotação, mas apresentou limitações em relação à sensibilidade a condições de iluminação e ruído na aquisição das imagens.

Ainda nesse contexto, [Chen et al. \(2012\)](#) propuseram um esquema de ponderação dinâmica para combinar múltiplas características visuais, incluindo cores, textura e forma. Em seguida, os autores atribuíram pesos adaptativos a cada característica de acordo com métricas de similaridade calculadas em tempo real. Essa estratégia melhorou a robustez do sistema a variações de iluminação e contraste. Contudo, o método apresentou limitações: a complexidade computacional é alta, dificultando a implementação em dispositivos com restrição de recursos; além disso, o ajuste dos pesos depende de regras e parâmetros fixos, o que pode comprometer a flexibilidade do sistema em diferentes cenários.

Com uma proposta metodológica distinta, [Caban, Rosebrock e Yoo \(2012\)](#) desenvolveram um método baseado em modelos de distribuição de forma invariantes à rotação, que capturam a geometria global de comprimidos ao computar distâncias entre o ponto central do comprimido e pontos igualmente espaçados ao longo do seu contorno. Além da forma, a abordagem também considera cor e impressão, por meio de técnicas baseadas em histogramas e mapas de bordas, atingindo uma acurácia de 91,13% na identificação de 568 comprimidos prescritos nos EUA. Entretanto, o método assume que as imagens sejam capturadas com vista superior completa do comprimido, o que pode limitar seu desempenho em cenários reais, com ângulos parciais ou oclusões. Ademais, todas as características são combinadas com pesos fixos, o que pode não ser ideal para todos os casos.

Com o intuito de superar parte dessas limitações, [Yu et al. \(2015\)](#) propuseram o descritor TSDS (*Two-Step Sampling Distance Sets*), projetado para capturar a geometria das impressões. O método realiza uma dupla amostragem: a primeira seleciona pontos uniformemente ao longo do contorno do comprimido, enquanto a segunda foca em regiões de alta curvatura, permitindo uma representação da distribuição espacial global da impressão. Essa abordagem obteve 90,46% de acurácia *top-1* (considerando apenas a primeira predição), e 97,16% de acurácia *top-5* (considerando as cinco predições mais prováveis), em um banco de dados com 2.500 comprimidos. No entanto, o sistema apresentou sensibilidade à escolha do parâmetro de curvatura, que afeta diretamente o desempenho.

Ampliando o foco na leitura de impressões, [Suntronsuk e Ratanotayanon \(2016, 2017\)](#) introduziram abordagens especializadas para extração automática de impressões gravadas em comprimidos, com ênfase na binarização de regiões das impressões para aplicação de reconhecimento óptico de caracteres (*optical character recognition*, OCR). Os métodos combinam técnicas de realce de bordas, filtros morfológicos e transformações geométricas para identificar e isolar a área do texto, mesmo em superfícies curvas e com baixo contraste. As imagens são então binarizadas usando limiares de Otsu ou segmentações baseadas em agrupamento *k*-médias, e processadas pelo motor OCR Tesseract ([SMITH, 2007](#)), previamente treinado com fontes específicas extraídas dos próprios dados.

As melhores configurações alcançaram *f-measure*² de até 77,30% para os textos impressos. Contudo, os métodos apresentaram limitações consideráveis na extração de gravações em baixo relevo, devido à baixa separabilidade entre plano de fundo e caracteres, além de dependência significativa da iluminação e necessidade de pré-treinamento do OCR para cada tipo de fonte.

2.2 Métodos estatísticos e aprendizado de máquina

A partir de 2015, há uma mudança significativa na literatura para técnicas baseadas em aprendizado de máquina e redes neurais. Chupawa e Kanjanawanishkul (2015) criaram um sistema de identificação de comprimidos baseado em uma rede neural artificial treinada com características extraídas exclusivamente das impressões dos comprimidos. O método utiliza uma técnica de amostragem radial para capturar a distribuição de pixels da impressão em diferentes regiões concêntricas do comprimido, gerando um vetor de características invariante à rotação. Esse vetor é então utilizado como entrada em uma rede perceptron com três camadas não lineares, capaz de classificar comprimidos visualmente semelhantes em termos de cor e forma. Apesar de alcançar 94,40% de acurácia em um conjunto composto por seis classes, o sistema apresentou limitações, como o uso de um número reduzido de classes, a dependência de pré-processamentos sensíveis à perda de detalhes da impressão e a necessidade de ajustes manuais de limiares e parâmetros de segmentação.

Na mesma direção, Ushizima *et al.* (2015) propuseram um arcabouço avaliativo para reconhecimento de comprimidos, utilizando o então recente conjunto de imagens da NLM, denominado NLM PIR. O trabalho introduziu algoritmos de segmentação e extração de atributos morfométricos, como circularidade, proporção e solidez, além de descritores estatísticos de intensidade, como perímetro, largura e altura. Esses descritores foram empregados para agrupar os comprimidos em classes com base na similaridade visual, por meio dos métodos *k*-médias e mapas auto-organizáveis.

O estudo concluiu que os descritores foram capazes de dividir as formas dos comprimidos em quatro classes: circular, cápsula, oval e irregular. No entanto, o sistema apresentou limitações, como a ausência de atributos textuais (isto é, impressões) na caracterização e a dificuldade de separação entre formas complexas com contornos irregulares. Além disso, o foco do trabalho restringiu-se às imagens de referência em alta resolução, não abordando as variações presentes nas imagens capturadas por usuários com *smartphones*.

Em continuidade a esses avanços, Vieira Neto *et al.* (2018) desenvolveram o CoforDes, um extrator de características invariante à rotação, que combina atributos de forma, como circularidade, proporção e solidez, com atributos de cor extraídos em torno do centro

² A *f-measure* (ou *f1-score*) é uma métrica de avaliação que combina precisão e revocação em um único valor, representando a média harmônica entre ambas. Valores mais altos indicam melhor equilíbrio entre acertos e abrangência na detecção correta das classes.

geométrico do comprimido, utilizando estatísticas no espaço de cor CIELAB. Os autores avaliaram o desempenho do CoforDes nos conjuntos NLM PIR e PILL BR (um conjunto de dados composto por comprimidos brasileiros), alcançando mais de 99,80% de acurácia e mostrando desempenho computacional compatível com aplicações em tempo real.

Adicionalmente, em uma análise comparativa, os autores verificaram que o CoforDes superou diversos extratores tradicionais, como *Gray-Level Co-occurrence Matrix* (GLCM) (HARALICK; SHANMUGAM; DINSTEIN, 1973), *Local Binary Patterns* (LBP) (OJALA; PIETIKÄINEN; MÄENPÄÄ, 2002), Momentos de Hu (HU, 1962) e Momentos de Zernike (TEAGUE, 1980). Apesar do desempenho relevante, o uso de imagens capturadas em condições parcialmente controladas permanece como limitação, uma vez que o extrator foi exposto a poucas condições reais, como desfoque, oclusão e variações de iluminação.

2.3 Redes neurais convolucionais

O advento do aprendizado profundo (*deep learning*), em particular das redes neurais convolucionais (*convolutional neural networks*, CNNs), revolucionou a pesquisa em reconhecimento de comprimidos, ao automatizar a extração de atributos e melhorar significativamente o desempenho dos modelos. Wong *et al.* (2017) foram pioneiros na utilização de CNNs para identificação automática de comprimidos, treinando modelos a partir de imagens capturadas por dispositivos móveis em ambientes não controlados. O método emprega etapas de correção geométrica e segmentação baseada em detecção de saliência para isolar o comprimido, seguido por uma arquitetura similar à AlexNet (KRIZHEVSKY; SUTSKEVER; HINTON, 2012), pré-treinada no ImageNet (DENG *et al.*, 2009) e ajustada com imagens reais e aumentadas de comprimidos de um conjunto de imagens próprio.

Nesse estudo, o método superou técnicas baseadas em atributos manuais, alcançando 95,35% de acurácia *top-1*, 98,75% de acurácia *top-5* e 99,55% de acurácia *top-10*, em um conjunto com 400 classes. Contudo, os autores destacaram as seguintes limitações: o número restrito de classes utilizadas, a sensibilidade a imagens desfocadas ou com pouca informação visual, e a necessidade de ampliação do conjunto de dados para garantir robustez em contextos clínicos reais.

Dando continuidade a essa linha de pesquisa, Zeng, Cao e Zhang (2017) desenvolveram o MobileDeepPill, um sistema de reconhecimento de comprimidos baseado em redes neurais profundas, projetado especificamente para operar com eficiência em dispositivos móveis. O sistema adota uma arquitetura com múltiplas CNNs para capturar conjuntamente as características de forma, cor e impressão dos comprimidos, utilizando uma função de custo triplo (*triplet loss*) para aumentar a robustez frente às variações de iluminação, orientação e qualidade das imagens capturadas por *smartphones*.

Para viabilizar a execução em hardware com recursos limitados, os autores propuseram uma técnica de compressão baseada em destilação de conhecimento (*knowledge distillation*), reduzindo em até 86,80% o número de parâmetros e em 62,80% o custo computacional, mantendo desempenho comparável ao modelo original. O sistema obteve 73,70% de acurácia *top-1* e 95,60% de acurácia *top-5*. Entretanto, o método apresentou limitações: o desempenho reduzido em imagens com múltiplos comprimidos, a dependência de segmentação precisa e acurácia inferior quando apenas uma face do comprimido está disponível (52,70% de acurácia *top-1*), além da necessidade de ajustes para generalização a novas classes.

Seguindo uma abordagem integrada de detecção e classificação, [Ou et al. \(2018\)](#) projetaram um sistema estruturado em duas etapas: localização e reconhecimento. A etapa de localização utiliza uma ResNet-50 ([HE et al., 2016](#)) como modelo base para construir mapas de características multiescala e prever as coordenadas dos objetos, enquanto a etapa de classificação aplica a arquitetura Xception ([CHOLLET, 2017](#)) sobre as regiões detectadas para identificar a classe do comprimido.

Avaliado em um conjunto com 131 classes e mais de 470 mil imagens, o sistema alcançou 79,40% de acurácia *top-1* e 91,80% de acurácia *top-5*. Apesar dos resultados expressivos, o modelo possui uma limitação importante: a ausência de avaliação em cenários com classes não vistas previamente, o que restringe sua aplicabilidade em contextos clínicos dinâmicos, nos quais novos medicamentos são frequentemente introduzidos.

De forma semelhante, [Delgado et al. \(2019\)](#) propuseram um sistema de identificação de medicamentos em imagens reais, baseado em um fluxo de processamento composto por duas etapas: segmentação e classificação. Inicialmente, os comprimidos são localizados por meio de uma rede totalmente convolucional (*fully convolutional network*) treinada com imagens sintéticas que simulam diferentes variações de fundo, iluminação e posicionamento. As regiões segmentadas são então normalizadas e encaminhadas a uma CNN treinada para a tarefa de classificação, cujo modelo base é escolhido entre diferentes arquiteturas, como ResNet-50, InceptionV3 ([SZEGEDY et al., 2016](#)), MobileNet ([HOWARD et al., 2017](#)) e SqueezeNet ([IANDOLA et al., 2016](#)).

Os autores realizaram uma comparação sistemática entre essas arquiteturas, avaliando a acurácia e o custo computacional de cada uma. O modelo baseado em ResNet-50 apresentou o melhor equilíbrio entre precisão e desempenho, alcançando 77% de acurácia *top-1* e 95,30% de acurácia *top-5* nas imagens de consulta do conjunto NLM PIR. Embora o sistema tenha se mostrado robusto em cenários controlados, sua generalização depende da diversidade do conjunto de treinamento, especialmente para capturas feitas em condições não padronizadas.

2.4 Aprendizado com poucos exemplos

Apesar dos avanços significativos proporcionados pelas CNNs, a necessidade de grandes conjuntos de dados rotulados para treinamento dos modelos permanece como um obstáculo relevante para a generalização dos sistemas de identificação de comprimidos. Essa limitação motivou o surgimento de modelos treinados com poucos exemplos disponíveis, fundamentados no aprendizado com poucos exemplos (*few-shot learning*).

Ling *et al.* (2020) foram pioneiros na aplicação dessa abordagem no reconhecimento de comprimidos. Eles propuseram uma arquitetura baseada em aprendizado métrico³ (*metric learning*), com otimização de custo triplo, composta por uma rede com múltiplos fluxos paralelos, cada um dedicado ao processamento de diferentes tipos de características visuais: cor, textura, contorno e impressão. Inicialmente, o treinamento é realizado com a estratégia *batch-all* e, posteriormente, refinado com exemplos mais difíceis utilizando a abordagem *batch-hard*. Para o fluxo de impressão, os autores treinaram o modelo Deep TextSpotter (BUSTA; NEUMANN; MATAS, 2017).

Avaliada nos conjuntos NLM PIR e CURE (proposto pelos autores), a abordagem superou métodos anteriores, como o MobileDeepPill, alcançando 68,30% de acurácia no conjunto CURE e 65,90% no conjunto NLM PIR. Embora os autores não discutam extensivamente as limitações, o método pressupõe a disponibilidade de segmentações precisas e pode ter o desempenho reduzido em casos onde as impressões não estão visíveis.

Na mesma direção, Usuyama *et al.* (2020) introduziram o ePillID, um conjunto de dados projetado para *benchmarking* de modelos de identificação de comprimidos treinados em regime de poucos exemplos. O conjunto é composto por mais de 13 mil imagens, organizadas em 8.184 classes aparentes (dois lados para 4.092 comprimidos). Para a maioria das classes há apenas uma imagem de referência, o que configura um cenário de reconhecimento desafiador. Os autores propuseram um *benchmarking* com protocolos de avaliação realistas, com 1 a 5 amostras por classe e métricas como acurácia (*top-1* e *top-5*), *mean average precision* (MAP) e *global average precision* (GAP).

No decorrer do estudo, os autores examinaram diversos modelos baseados em classificação e aprendizado métrico, sendo que redes neurais profundas combinadas a um agrupamento bilinear compacto (*compact bilinear pooling*) e aprendizado métrico com múltiplas cabeças (*multi-head metric learning*) obtiveram os melhores resultados. Apesar do alto desempenho em cenários com imagens de ambos os lados do comprimido (MAP > 95%), os modelos ainda falharam na distinção de classes muito semelhantes, e o *benchmarking* evidenciou dificuldades com impressões pouco visíveis, superfícies reflexivas e escassez de imagens capturadas por usuários para todas as classes.

³ Aprendizado métrico é uma abordagem de aprendizado de máquina que busca aprender uma função de distância ou similaridade entre amostras, de modo que amostras da mesma classe fiquem próximas no espaço de representação, enquanto amostras de classes distintas fiquem afastadas.

Em um avanço mais recente, [Zhang et al. \(2024\)](#) projetaram o DBC-FSCIL, um arcabouço incremental e compatível com aprendizado com poucos exemplos (*few-shot class-incremental learning*) voltado à identificação contínua de novos comprimidos. O sistema combina duas estratégias: aprendizado com compatibilidade futura, baseado em geração de classes virtuais e na função de custo centro-tríplo (*center-triplet loss*) para melhorar a separação entre classes; e aprendizado com compatibilidade retroativa, utilizando destilação de conhecimento e síntese de pseudo-características com incerteza controlada. A abordagem permitiu adicionar novas classes sem esquecer as antigas e sem armazenar imagens reais.

Os autores avaliaram o DBC-FSCIL nos conjuntos FCPill (proposto por eles próprios) e CURE, e o modelo superou métodos do estado da arte, alcançando alta acurácia incremental⁴. Embora o modelo tenha apresentado resultados superiores em cenários incrementais, os autores não discutiram explicitamente limitações relacionadas à complexidade computacional, à segmentação em imagens reais ou à necessidade de adaptação em ambientes clínicos distintos.

Por outro lado, [Chee et al. \(2024a\)](#) investigaram o uso de transferência de aprendizado (*transfer learning*) para reconhecimento de comprimidos com poucos exemplos, propondo uma arquitetura baseada em redes prototípicas (*prototypical networks*) com ResNet-101 ([HE et al., 2016](#)) pré-treinada no ImageNet como modelo base. A rede foi completamente ajustada durante o treinamento em regime de *1-shot 5-way*, buscando maximizar a generalização a partir de conjuntos limitados de imagens.

A abordagem alcançou acurácia de 89,65% no conjunto CURE, superando métodos como *Category Traversal Module* (CTM) ([LI et al., 2019](#)), *Meta-Transfer Learning* (MTL) ([SUN et al., 2019](#)) e *Multi-Stream* (MS) ([LING et al., 2020](#)) em mais de 35 pontos percentuais. A integração do conhecimento prévio extraído do ImageNet acelerou a convergência e melhorou a formação dos protótipos. Embora o artigo não discuta limitações diretamente, a abordagem pressupõe a disponibilidade de pré-treinamento em domínios amplos e não avalia a robustez frente a variações de qualidade ou mudanças de domínio.

2.5 Arquiteturas com múltiplos fluxos e comitês de classificadores

A fim de superar limitações presentes em abordagens tradicionais e explorar de forma mais abrangente as diferentes características visuais dos comprimidos, alguns autores desenvolveram arquiteturas que processam múltiplos tipos de atributos em fluxos paralelos. [Rádli, Vörösházi e Czúni \(2023\)](#) desenvolveram um modelo de reconhecimento de comprimidos semelhante ao de [Ling et al. \(2020\)](#), com a inclusão de um fluxo dedicado

⁴ Alta acurácia incremental é a capacidade de um modelo de aprendizado de máquina manter ou melhorar sua acurácia à medida que novos dados ou classes são adicionados ao treinamento, sem esquecer o que já foi aprendido.

para LBP. As saídas desses fluxos são combinadas por meio de mecanismos de autoatenção, formando um vetor de representação (*embedding*) métrico robusto.

Utilizando EfficientNet-B0 (TAN; LE, 2019) como modelo base e função de custo triplo com mineração de trios (*triplet loss mining*), o modelo alcançou 95,30% de acurácia no conjunto OGYEI (proposto pelos autores) e superou o YOLOv7 (WANG; BOCHKOVSKIY; LIAO, 2023) em condições controladas. Embora os resultados sejam promissores, o sistema foi avaliado apenas em imagens bem segmentadas e não incorporou informações textuais extraídas por OCR, como impressões gravadas, o que pode restringir sua capacidade de distinguir comprimidos com aparência visual muito semelhante.

Em outra linha de pesquisa, Nguyen *et al.* (2022) introduziram um arcabouço de reconhecimento contextual que integra representações visuais de comprimidos com informações relacionais obtidas por meio de grafos de conhecimento médico (*medical knowledge graphs*). A arquitetura combina vetores extraídos de imagens de comprimidos com vetores gerados por redes convolucionais em grafos, capturando relações implícitas de coocorrência entre medicamentos frequentemente prescritos em conjunto. O modelo obteve ganhos de até 34,10% em *f-measure* em relação a métodos puramente visuais. Embora os autores não discutam as limitações do modelo explicitamente, a abordagem depende de dados médicos estruturados de alta qualidade para a construção do grafo, e sua eficácia pode ser reduzida em contextos com grafos ruidosos ou incompletos.

Por sua vez, Chee *et al.* (2024b) investigaram comitês de classificadores (*ensembles*) para aprendizado com poucos exemplos na identificação de comprimidos, comparando votações majoritárias, médias de probabilidades e empilhamento (*stacking*). O estudo concluiu que o empilhamento de comitês com meta-classificadores apresentou desempenho superior em conjuntos desbalanceados, ao aprenderem padrões de erro específicos dos modelos base. Entretanto, o método requer maior capacidade de treinamento supervisionado e sofre com sobreajuste em classes com poucos exemplos, sendo sensível à qualidade dos vetores de representação individuais utilizados como entrada no meta-modelo.

2.6 Detecção e reconhecimento simultâneos

Tan *et al.* (2021a, 2021b) conduziram uma análise comparativa entre arquiteturas de detecção de objetos, como RetinaNet (LIN *et al.*, 2020), SSD (LIU *et al.*, 2016), *Faster R-CNN* (REN *et al.*, 2015) e YOLOv3 (REDMON; FARHADI, 2018), aplicadas à identificação de comprimidos em tempo real. Os autores avaliaram os modelos utilizando as métricas MAP, quadros por segundo (*frames per second*, FPS) e desempenho em amostras difíceis. Embora o RetinaNet tenha atingido a maior MAP (82,89%), o YOLOv3 demonstrou superioridade em FPS, alcançando desempenho em tempo real com 51 FPS e mantendo MAP competitivo (80,69%). O SSD teve desempenho inferior nos dois aspectos.

Os autores concluíram que o YOLOv3 oferece o melhor equilíbrio entre precisão e velocidade, sendo adequado para ambientes farmacêuticos com alta demanda. Apesar dos resultados satisfatórios, os modelos enfrentaram dificuldades em detectar comprimidos com tamanhos reduzidos, formas semelhantes ou oclusões parciais, além de o desempenho poder ser comprometido em ambientes fora do controle experimental. Essas limitações reforçam a necessidade de testar essas abordagens em cenários clínicos mais desafiadores, com variabilidade de iluminação, ruído e múltiplos comprimidos.

Em continuidade a essa linha de investigação, [Nguyen et al. \(2023\)](#) propuseram o PGPNet, um arcabouço para detecção e reconhecimento de múltiplos comprimidos em imagens capturadas em condições não controladas. O sistema utiliza redes neurais em grafos para incorporar informações relacionais entre comprimidos com base em três grafos: coocorrência em prescrições, proporção de tamanhos e correlação visual semântica. Essas informações são fundidas aos atributos visuais extraídos por uma *Faster R-CNN*, permitindo lidar com amostras com aparência semelhante.

O modelo superou os desempenhos de modelos baseados em YOLOv5 ([JOCHER; CHAURASIA, 2022](#)) e *Faster R-CNN*, melhorando a MAP no conjunto de dados COCO ([LIN et al., 2014](#)) em até 12%. Apesar dos avanços, a abordagem apresentou limitações, como o custo computacional elevado e a dependência de segmentação precisa, além da necessidade de grafos baseados em dados externos (prescrições), nem sempre disponíveis em ambientes clínicos. Os autores também reconheceram a necessidade de incorporar relações espaciais explícitas entre comprimidos e explorar estratégias que reduzam a complexidade e melhorem a generalização do modelo em cenários com dados ruidosos ou ausentes.

2.7 Contribuições teóricas e interdomínios

[Ozmermer et al. \(2021\)](#) apresentaram o sistema VeriMedi, que combina segmentação via *Mask R-CNN* ([HE et al., 2017](#)) e o treinamento de um modelo de aprendizado métrico profundo através da *proxy anchor loss*, visando reconhecer comprimidos em frascos fechados por meio de imagens obtidas por um dispositivo de IoT (*Internet of Things*). A introdução de inicialização otimizada para os vetores de proxy permitiu maior separabilidade entre classes semelhantes no espaço de representação, favorecendo tarefas de reconhecimento contínuo. No entanto, o sistema dependeu de segmentações precisas, foi sensível a ruídos visuais introduzidos por embalagens translúcidas e enfrentou desafios em ambientes com múltiplas instâncias parcialmente ocultas.

Em continuidade às propostas voltadas à adaptação incremental, [Nguyen et al. \(2022\)](#) propuseram o arcabouço CG-IMIF, um modelo incremental baseado em múltiplos fluxos com fusão intermediária, que incorpora informações de cor (histogramas RGB) como orientação auxiliar para minimizar o esquecimento catastrófico (*catastrophic forget-*

ting). A abordagem combina fluxos paralelos para imagens RGB e atributos cromáticos, integrando-os com mecanismos de destilação de conhecimento em um cenário de aprendizado incremental. Avaliado no conjunto VAIPE-PCIL, o modelo superou métodos anteriores em acurácia e estabilidade. Entretanto, ele requer armazenamento de exemplos antigos e não explora atributos textuais ou estruturais, o que pode limitar sua robustez em dados de baixa qualidade ou comprimidos com aparência ambígua.

Por outro lado, [Chang et al. \(2020\)](#) desenvolveram o MedGlasses, um sistema vestível baseado em óculos inteligentes com detecção local embarcada, voltado para pacientes com deficiência visual. O sistema utiliza os métodos SSD e ResNet-50 para localizar e classificar comprimidos em tempo real, além de emitir alertas sonoros sobre a correção da medicação. A arquitetura inclui um módulo de reconhecimento inteligente, um aplicativo móvel e uma plataforma de gerenciamento em nuvem. Embora tenha atingido acurácia de 95,10% em testes controlados, o sistema é limitado pela dependência de hardware específico, pela sensibilidade à iluminação e à posição da câmera, e pela falta de adaptação automática a novos medicamentos não registrados previamente.

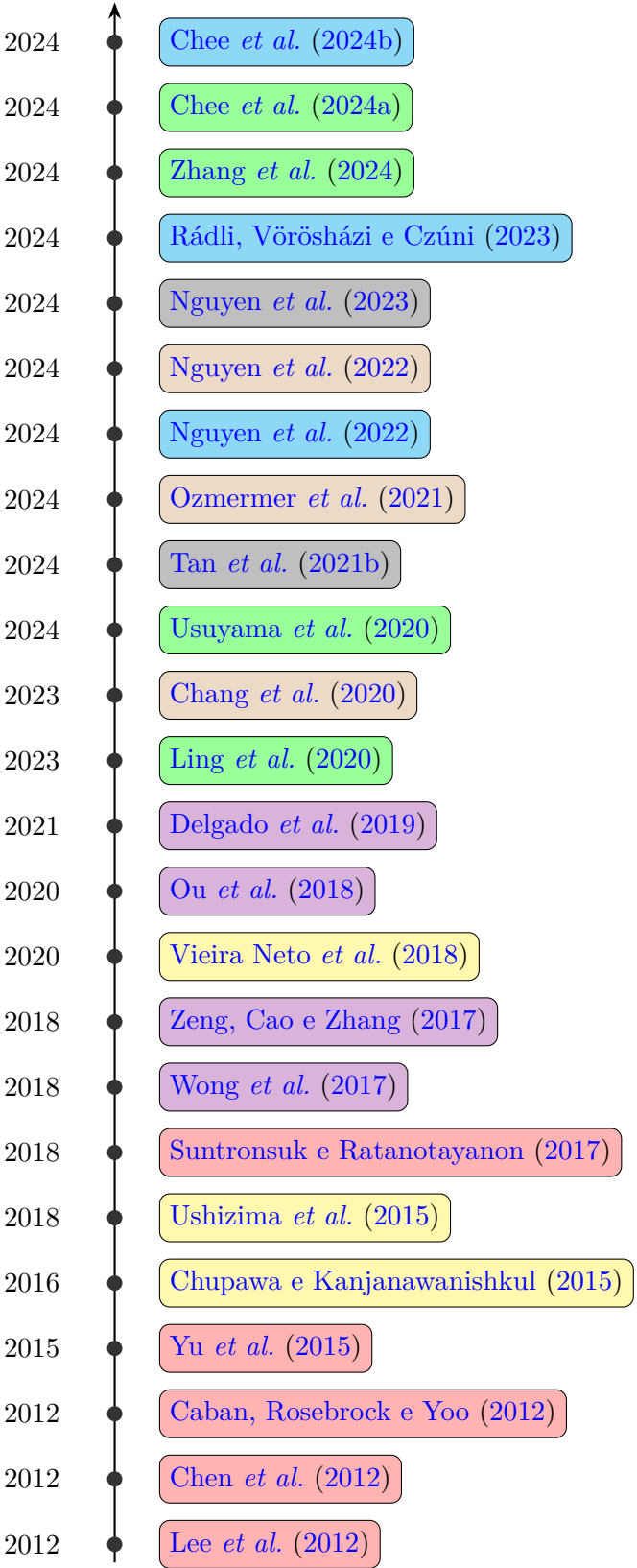
2.8 Considerações finais

Ao longo da última década, a literatura sobre identificação de comprimidos evoluiu de abordagens baseadas em descritores manuais para sistemas baseados em aprendizado profundo, capazes de extrair representações visuais complexas de forma automática. Inicialmente, os métodos concentraram-se em atributos geométricos e textuais extraídos de imagens de referência, alcançando resultados satisfatórios em ambientes controlados, mas com baixa generalização para imagens de consulta (capturadas por usuários). Com o advento das CNNs, tornou-se possível integrar informações de cor, forma e impressão em fluxos de processamento únicos de detecção e classificação, ampliando significativamente a acurácia e a robustez.

Posteriormente, surgiram arquiteturas voltadas à execução em dispositivos móveis, ao aprendizado com poucos exemplos e à adaptação incremental, visando reduzir a dependência de grandes volumes de dados rotulados e melhorar a escalabilidade dos sistemas em contextos clínicos reais. Esses avanços consolidaram uma transição para modelos mais autônomos, eficientes e generalizáveis, ao mesmo tempo em que evidenciaram desafios persistentes, como a variabilidade das imagens de consulta, a escassez de dados anotados e a necessidade de mecanismos de adaptação contínua.

A Figura 2 apresenta uma linha do tempo cronológica com os trabalhos da literatura sobre identificação de comprimidos que discutimos neste capítulo. Ela ilustra a transição das abordagens tradicionais de visão computacional para técnicas baseadas em aprendizado profundo, incluindo métodos mais recentes de aprendizado com poucos exemplos.

Figura 2 – Linha do tempo da literatura acadêmica sobre identificação de comprimidos. Legenda de cores: ■ visão computacional tradicional, ■ transição para aprendizado de máquina, ■ CNN clássicas, ■ aprendizado com poucos exemplos, ■ detecção, ■ métodos híbridos, ■ dados multimodais.



Fonte: Autor.

3 Conjuntos de dados

O desenvolvimento de sistemas de identificação automática de comprimidos por imagem requer conjuntos de dados visualmente representativos, bem rotulados e que reflitam os desafios encontrados em ambientes clínicos e farmacêuticos. Neste capítulo, descrevemos dois conjuntos amplamente utilizados em tarefas de identificação de comprimidos. O primeiro, denominado NLM PIR, é composto por imagens capturadas em condições controladas e não controladas, oferecendo um cenário adequado para avaliação de modelos de classificação. O segundo, denominado CURE, foi construído com foco em problemas com poucas amostras por classe, reunindo imagens com grande diversidade visual e condições de captura variadas. A seguir, detalhamos as características, a estrutura e a relevância de cada conjunto para as propostas desta dissertação.

3.1 NLM PIR

Em 2016, a NLM promoveu um desafio público denominado *Pill Image Recognition Challenge* (PIR) ([MEDICINE, 2019](#)), com o objetivo de incentivar o desenvolvimento de algoritmos e sistemas automatizados capazes de reconhecer comprimidos a partir de imagens capturadas em condições não controladas. Como parte da iniciativa, ela disponibilizou um conjunto de dados específico para essa tarefa, denominado NLM PIR. Esse conjunto contém 7 mil imagens de mil comprimidos distintos, todos aprovados pela FDA. As imagens foram coletadas com o intuito de representar a diversidade visual presente nos medicamentos comerciais, abrangendo variações de forma, cor, tamanho e textos impressos (*imprints*).

Os comprimidos presentes na base foram selecionados a partir da coleção RxIMAGE⁵, um repositório mantido pela FDA que reúne imagens de alta qualidade de medicamentos aprovados para uso humano. A base de dados NLM PIR foi projetada para servir como referência em pesquisas sobre identificação visual de medicamentos, refletindo os desafios encontrados em cenários do mundo real.

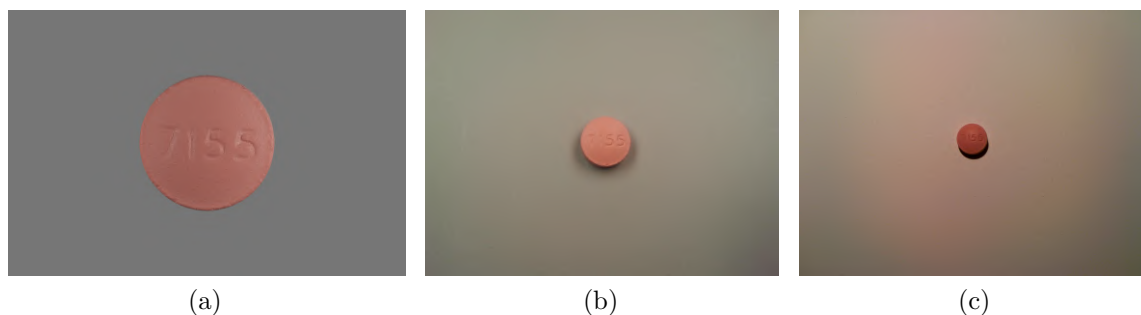
Cada comprimido é representado por duas imagens de referência, uma da frente e outra do verso, e cinco imagens de consulta. As imagens de referência foram obtidas sob condições controladas, garantindo que as características principais de cada comprimido fossem claramente visíveis e consistentes. Para os comprimidos em formato de cápsula, o lado selecionado corresponde àquele em que o texto impresso é integralmente visível, capturado com uma visão perpendicular ao plano da impressão. Essa escolha assegura que

⁵ A coleção RxIMAGE é um produto resultante do projeto *Computational Photography Project for Pill Identification* (C3PI), também realizado pela NLM.

informações críticas, como a forma e o texto impresso, estejam claramente destacadas nas imagens de referência, reduzindo ambiguidades no processo de reconhecimento.

As imagens de consulta, por sua vez, refletem variações típicas de cenários reais, como diferenças nos ângulos de visão e nas condições de iluminação. Essa variabilidade é essencial para avaliar a capacidade dos modelos de reconhecer os comprimidos, mesmo sob condições adversas. A combinação entre imagens de referência controladas e imagens de consulta mais diversificadas cria um cenário desafiador e realista para o desenvolvimento de sistemas automáticos de identificação de comprimidos. A Figura 3 ilustra exemplos de imagens de referência e de consulta para um mesmo comprimido.

Figura 3 – Exemplos de imagens de referência e de consulta para um mesmo comprimido do conjunto de dados NLM PIR. Da esquerda para a direita: imagem de referência (verso do comprimido), imagem de consulta 1 e imagem de consulta 2. Os nomes originais das amostras estão listados no Apêndice D.



Fonte: [Medicine \(2019\)](#).

O conjunto de dados NLM PIR também possui uma ampla variedade visual, especialmente em relação às formas geométricas e às combinações cromáticas. As formas incluem estruturas simétricas e regulares, como círculos e ovais, bem como formatos mais complexos e assimétricos. As cores variam entre tons claros e escuros, abrangendo comprimidos monocromáticos e bicolores, o que reforça a heterogeneidade visual do conjunto. A Figura 4 apresenta um mosaico com 12 imagens de referência do conjunto NLM PIR, organizadas em uma grade de três linhas e quatro colunas.

3.2 CURE

O conjunto de dados CURE foi criado por [Ling et al. \(2020\)](#) com o propósito de avançar a pesquisa em identificação automática de comprimidos sob o paradigma de aprendizagem com poucos exemplos. Diferentemente de outros conjuntos de dados, o CURE foi construído para refletir cenários do mundo real nos quais há uma quantidade limitada de amostras rotuladas por classe, situação comum em ambientes clínicos e farmacêuticos.

Esse conjunto contém 8.973 imagens de 196 categorias distintas de comprimidos, com cerca de 45 amostras por classe. As imagens foram obtidas a partir de quatro

Figura 4 – Mosaico com 12 imagens de referência do conjunto de dados NLM PIR. Os nomes originais das amostras estão listados no Apêndice D.



Fonte: [Medicine \(2019\)](#).

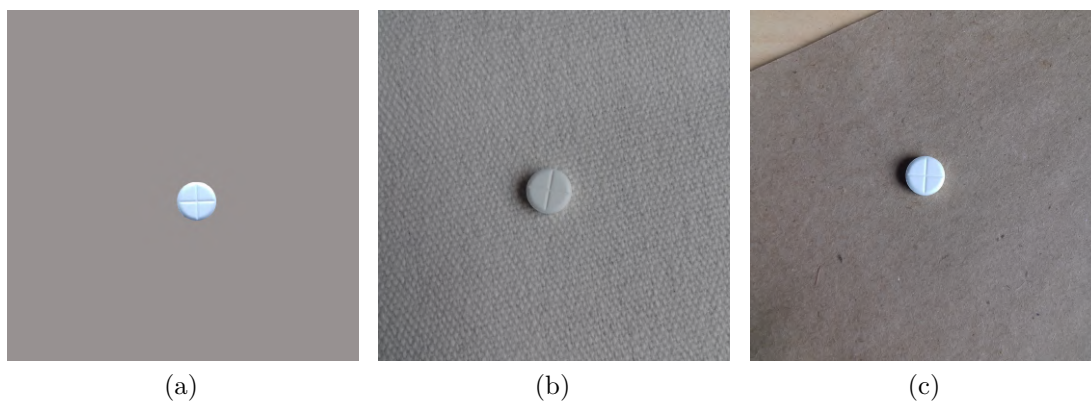
dispositivos: três *smartphones* (Samsung SM-J320FN, SM-N920S e LG F500L) e um sistema dedicado denominado *Multi-Pill Identifier* (MPI), composto por uma câmera e um Raspberry Pi montados sobre uma estrutura de suporte. O MPI foi projetado para uso em ambientes domésticos e farmácias, permitindo a digitalização de comprimidos a fim de auxiliar na sua identificação e verificar possíveis interações medicamentosas.

Assim como no conjunto NLM PIR, o CURE possui imagens de referência e de consulta para cada comprimido. Especificamente, as imagens de consulta foram capturadas em condições reais, apresentando ampla variabilidade de fundo, iluminação, foco e orientação. Para cada comprimido, foram consideradas três condições de iluminação (luz interna, luz externa fraca e luz externa intensa) e múltiplos níveis de textura e zoom. As imagens de referência, por sua vez, foram geradas a partir das melhores imagens de consulta disponíveis. As regiões correspondentes aos comprimidos foram segmentadas manualmente e inseridas em fundos neutros (cinza), criando uma representação limpa e padronizada. A Figura 5 ilustra exemplos de imagens de referência e de consulta para um mesmo comprimido.

A Figura 6 apresenta uma amostra visual da diversidade de comprimidos presentes no conjunto CURE. As 12 imagens ilustram variações significativas de cor, forma, plano de fundo e orientação, refletindo os desafios reais enfrentados por sistemas automatizados de identificação. É possível observar a presença de comprimidos com formas diferentes, bem como casos com baixa nitidez ou iluminação desfavorável, evidenciando a importância de estratégias e algoritmos robustos e generalizáveis.

Uma característica importante do CURE é a presença de amostras particularmente desafiadoras, denominados *hard samples*. Elas se dividem em dois tipos principais:

Figura 5 – Exemplos de imagens de referência e de consulta para um mesmo comprimido do conjunto de dados CURE. Da esquerda para a direita: imagem de referência (verso do comprimido), imagem de consulta 1 e imagem de consulta 2. Os nomes originais das amostras estão listados no Apêndice D.



Fonte: Ling *et al.* (2020).

Figura 6 – Mosaico com 12 imagens de consulta do conjunto de dados CURE. Os nomes originais das amostras estão listados no Apêndice D.



Fonte: Ling *et al.* (2020).

- a) **Positivos difíceis** (*hard positives*) são imagens do mesmo comprimido com diferenças significativas de aparência devido à iluminação, foco, ruído etc;
- b) **Negativos difíceis** (*hard negatives*) são comprimidos de classes distintas que compartilham características visuais como forma e cor.

Essas amostras refletem a realidade de ambientes não controlados, nos quais os sistemas de reconhecimento devem ser robustos à variabilidade intra e entre classes. A Figura 7 ilustra exemplos de ambos os tipos de amostras difíceis.

Figura 7 – Exemplos de amostras difíceis do conjunto de dados CURE. A primeira linha ilustra positivos difíceis. A segunda linha apresenta negativos difíceis. Os nomes originais das amostras estão listados no Apêndice D.



Fonte: [Ling et al. \(2020\)](#).

Em comparação com a base NLM PIR, amplamente utilizada no desafio promovido pela NLM em 2016, o CURE se destaca por apresentar maior diversidade nas condições de captura e quantidade de amostras por classe. A Tabela 1 resume as principais diferenças entre os dois conjuntos.

Tabela 1 – Características dos conjuntos de dados NLM PIR e CURE.

Característica	Conjunto de dados	
	NLM PIR	CURE
Número de imagens	7.000	8.973
Número de classes	1.000	196
Número de amostras por classe	7	[40, 50]
Número de planos de fundo	1	6
Condições de iluminação	1	3

Fonte: [Ling et al. \(2020\)](#).

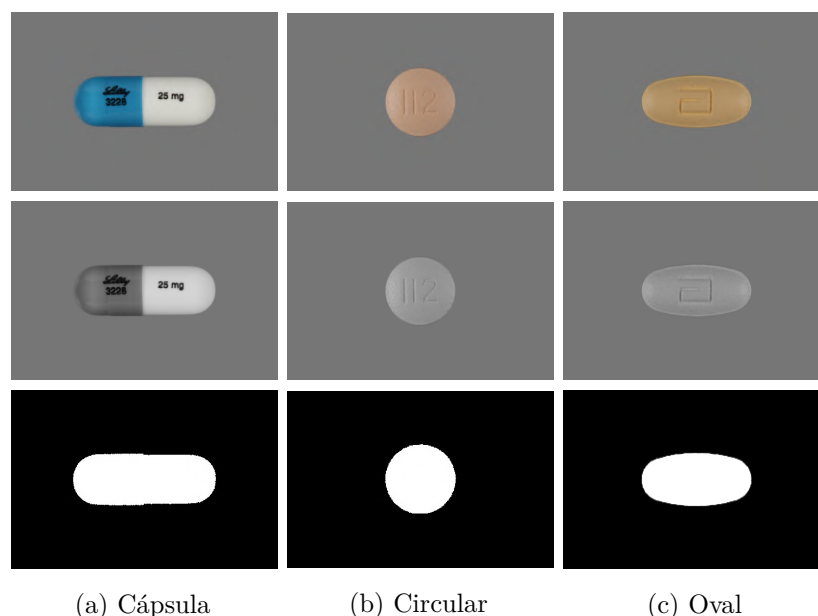
3.3 Uma estratégia preliminar para identificação de comprimidos

Nesta seção, propomos uma estratégia preliminar para a identificação de comprimidos em imagens capturadas em condições controladas. Para isso, adotamos técnicas tradicionais de processamento de imagem, por se tratarem de métodos de baixo custo computacional e boa interpretabilidade, adequados para cenários em que há poucos recursos computacionais disponíveis. A proposta é composta por duas etapas: segmentação baseada em limiarização e extração de atributos geométricos simples. Essa abordagem fornece uma base sólida para o desenvolvimento de sistemas mais avançados e robustos, que poderão ser aprimorados com o tempo, incorporando técnicas mais complexas e modelos de aprendizado de máquina mais sofisticados.

3.3.1 Segmentação

Para obter a região de um comprimido, primeiro transformamos o espaço de cores da imagem de entrada, do RGB para níveis de cinza. Conhecendo as condições de obtenção das imagens, podemos selecionar o pixel da coordenada $(0,0)$ e utilizá-lo como representante do plano de fundo da imagem. Isto posto, definimos com valor 0 todos os pixels cujos valores são iguais ao do pixel do plano de fundo; aos pixels cujos valores são diferentes, atribuímos o valor 1. A Figura 8 ilustra as subetapas do processo de segmentação em imagens do conjunto de dados NLM PIR.

Figura 8 – Ilustração das subetapas do processo de segmentação em imagens do conjunto de dados NLM PIR. Na primeira linha, estão as imagens de entrada. Na segunda linha, estão as imagens em níveis de cinza. Na terceira linha, estão as imagens segmentadas. Os nomes originais das amostras estão listados no Apêndice D.



Fonte: Autor.

3.3.2 Extração de atributos

A etapa de segmentação nos fornece regiões completas e devidamente delimitadas. Deste modo, podemos utilizar medidas geométricas simples para derivar atributos descritores de região e fronteira capazes de representar as formas dos comprimidos. Para entender com mais profundidade o significado dos atributos, definimos abaixo algumas medidas simples de uma região, a saber: área, fecho convexo, área do fecho convexo, eixo maior, eixo menor e perímetro (COSTA; CESAR JUNIOR, 2009; GONZALEZ; WOODS, 2018).

Área (A): número de pixels na região.

Fecho convexo: menor polígono convexo que contém todos os pixels da região.

Área do fecho convexo (AFC): número de pixels do fecho convexo da região.

Eixo maior (EM): segmento de linha que conecta os dois pontos extremos que compõem o diâmetro da fronteira da região.

Eixo menor (Em): segmento de linha perpendicular ao eixo maior e de comprimento tal que uma caixa que passa pelos quatro pontos exteriores de interseção da fronteira com os dois eixos envolve completamente a fronteira da região.

Perímetro (P): tamanho (comprimento) da fronteira de uma região.

A partir dessas medidas, nós derivamos os seguintes atributos: alongamento, circularidade, compacidade, razão perímetro-área, redondeza e solidez. A Tabela 2 apresenta um resumo dos atributos, com a definição de suas respectivas equações.

Tabela 2 – Atributos de forma.

Atributo	Equação
Alongamento (AG)	EM/Em
Circularidade (CL)	$(4\pi \cdot A)/P^2$
Compacidade (CP)	P^2/A
Razão perímetro-área (PA)	P/A
Redondeza (RD)	$(4 \cdot A)/(\pi \cdot EM^2)$
Solidez (SL)	A/AFC

Fonte: Costa e Cesar Junior (2009), Gonzalez e Woods (2018).

Particularmente:

- O alongamento é uma medida que expressa a proporção entre os eixos maior e menor. Um valor de alongamento igual a 1 indica que a forma é simétrica nos dois eixos, como um círculo ou quadrado, pois os eixos maior e menor têm

o mesmo comprimento. Alongamentos maiores que 1 indicam que a forma é mais alongada, como um retângulo ou uma elipse. Quanto maior o valor, mais alongada é a forma.

- b) A circularidade define quão próximo de um círculo perfeito uma forma está. Essa medida varia de 0 a 1, sendo que 1 indica uma forma perfeitamente circular. Quanto mais irregular ou alongada for a forma, menor será sua circularidade.
- c) A compacidade define o quão compacta é uma forma. Essa medida é inversamente proporcional à circularidade, sendo que valores menores indicam formas mais compactas e circulares. Para um círculo perfeito, o valor é 4π , e para um quadrado, 16.
- d) A razão perímetro/área relaciona a complexidade da fronteira de uma forma com a área que ela circunscreve. Formas com fronteiras mais complexas ou irregulares terão uma razão maior, enquanto formas mais simples e suaves terão uma razão menor.
- e) A redondeza é semelhante à circularidade. Contudo, possui um fator de normalização que a torna insensível a irregularidades nas bordas da forma.
- f) A solidez mede a densidade de uma forma. Vista sob a ótica da convexidade, a solidez representa o grau de convexidade de uma forma. Formas completamente sólidas (ou completamente convexas) possuem valor 1, enquanto formas com bordas irregulares ou contendo buracos possuem valor menor que 1.

Para fins de exemplificação, apresentamos na Tabela 3 os valores dos atributos de forma para as imagens binárias da Figura 8.

Tabela 3 – Valores dos atributos de forma para as imagens binárias na Figura 8.

Forma	AG	CL	CP	PA	RD	SL
Cápsula	2,7978	0,4952	25,3762	0,0069	0,3512	0,9841
Circular	1,0100	0,6518	19,2797	0,0079	0,9899	0,9839
Oval	1,9529	0,6015	20,8927	0,0074	0,5118	0,9868

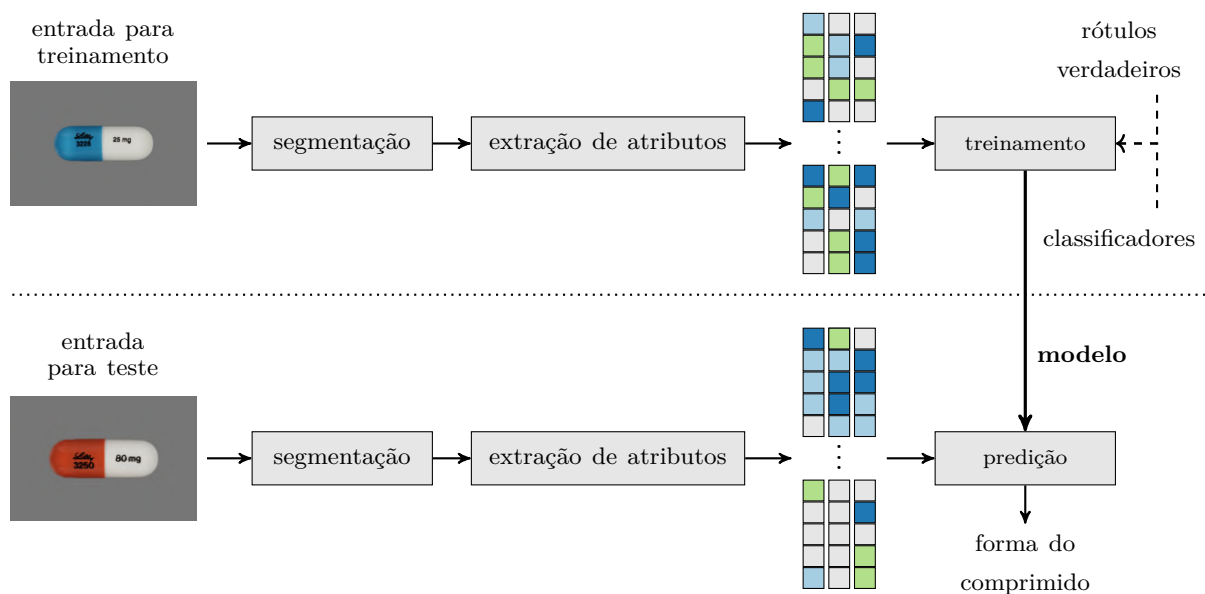
Fonte: Autor.

Esses atributos são capazes de descrever propriedades morfológicas relevantes para diferenciação de formas, como simetria, convexidade e complexidade da borda. Tais características são frequentemente utilizadas na literatura para análise de objetos binários (COSTA; CESAR JUNIOR, 2009; GONZALEZ; WOODS, 2018). Além disso, tais atributos são independentes da orientação ou escala das imagens e necessitam de baixos recursos computacionais para seus cálculos. Portanto, isso os torna adequados para aplicações que envolvem variações moderadas nas condições de aquisição das imagens.

4 Identificação de formas de comprimidos com técnicas de processamento de imagens

Neste capítulo, apresentamos a primeira contribuição desta dissertação, que consiste em uma metodologia voltada à identificação de formas de comprimidos por meio de técnicas tradicionais de processamento de imagens. O objetivo é avaliar o potencial de atributos geométricos simples na representação morfológica dos comprimidos e na discriminação entre classes de forma. Essa metodologia é composta por três etapas: segmentação, extração de atributos e classificação. As duas primeiras etapas seguem as definições apresentadas na Seção 3.3, enquanto a terceira etapa compreende a aplicação de diferentes classificadores supervisionados para identificar a forma predominante em cada imagem. A Figura 9 ilustra o fluxo de processamento adotado nesta proposta.

Figura 9 – Fluxo de processamento para identificação de formas de comprimidos.



Fonte: Autor.

Este capítulo está organizado da seguinte forma. Inicialmente, realizamos uma análise prévia das formas dos comprimidos, que nos permite simplificar o problema de classificação. Na sequência, discutimos o projeto dos classificadores, abordando aspectos como a preparação dos dados, os critérios de ajuste dos parâmetros, a seleção e o treinamento dos modelos. Os classificadores que utilizamos nessa etapa incluem algoritmos baseados em distância, probabilidade e funções, especificamente:

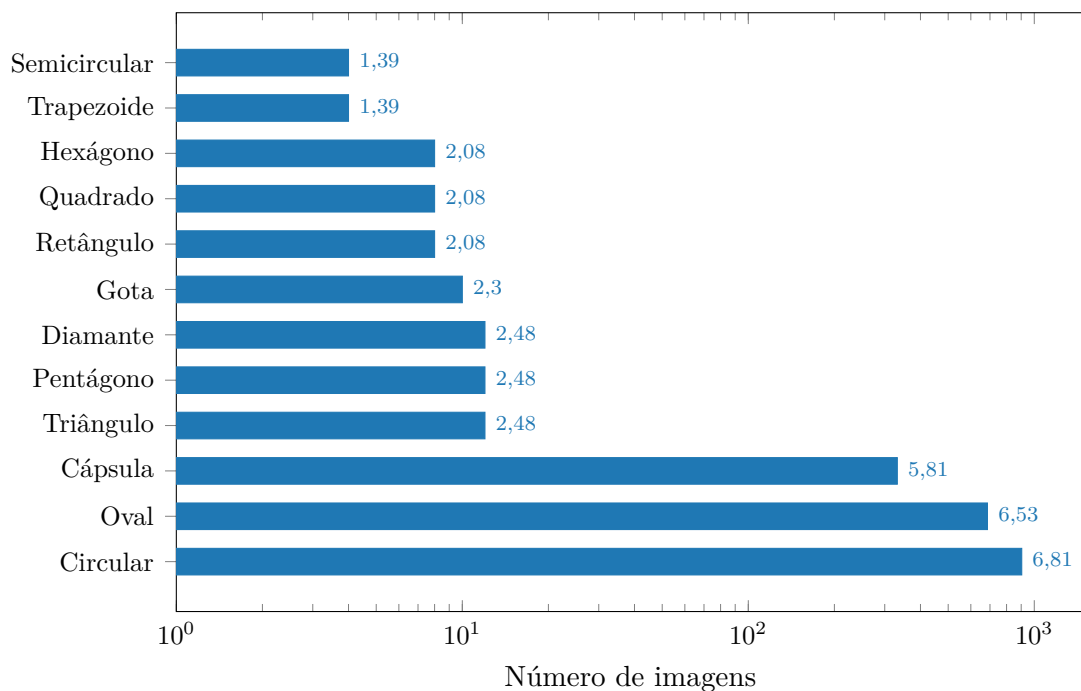
- a) *k*-Nearest Neighbors (*k*-NN) (COVER; HART, 1967);
- b) *Linear Discriminant Analysis* (LDA) (FISHER, 1936);
- c) *Quadratic Discriminant Analysis* (QDA) (ANDERSON, 1958);
- d) *Naïve Bayes* (NB) (DUDA; HART; STORK, 2001);
- e) *Multilayer Perceptron* (MLP) (RUMELHART; HINTON; WILLIAMS, 1986);
- f) *Extreme Learning Machine* (ELM) (HUANG; ZHU; SIEW, 2006);
- g) *Support Vector Machine* (SVM) (CORTES; VAPNIK, 1995).

O Apêndice A contém a descrição detalhada de cada classificador. Em seguida, apresentamos os grupos de experimentos realizados e discutimos os resultados obtidos. Por fim, apresentamos as conclusões e considerações finais sobre a proposta.

4.1 Análise das formas dos comprimidos

Antes de realizar a segmentação das imagens e a extração de atributos, conduzimos uma análise das formas dos comprimidos presentes no conjunto de dados NLM PIR. O objetivo dessa análise é compreender a distribuição das classes e antecipar possíveis desafios para as etapas posteriores de processamento. A Figura 10 apresenta a distribuição das formas dos comprimidos desse conjunto.

Figura 10 – Distribuição das formas dos comprimidos do conjunto de dados NLM PIR. Para uma melhor visualização, o eixo horizontal está em escala logarítmica.



Fonte: Autor.

Ao analisar o gráfico, identificamos um desbalanceamento significativo entre as classes. Para mitigar potenciais problemas no aprendizado dos classificadores, optamos por agrupar as classes com menor representação em uma classe denominada “Outra”. Especificamente, agrupamos as classes “Diamante”, “Gota”, “Hexágono”, “Pentágono”, “Quadrado”, “Retângulo”, “Semicircular”, “Trapezoide” e “Triângulo”. Embora a distribuição permaneça desbalanceada após esse agrupamento, conseguimos reduzir o número de classes para quatro: “Cápsula”, “Oval”, “Circular” e “Outra”. Essa estratégia deve diminuir a complexidade do espaço de características e favorecer a aprendizagem dos classificadores.

4.2 Experimentos e discussões

Nesta seção, descrevemos os experimentos e discutimos os resultados obtidos a partir da aplicação dos atributos de forma propostos na tarefa de identificação de formas de comprimidos. Os experimentos conduzidos visam avaliar a capacidade discriminativa dos atributos derivados de medidas geométricas simples, bem como o desempenho de diferentes algoritmos de classificação.

Os experimentos estão organizados em três grupos principais. Para o primeiro grupo, focamos em analisar o desempenho geral dos classificadores. Para isso, utilizamos as métricas de acurácia, precisão, revocação e o coeficiente de correlação de Matthews (*Matthews Correlation Coefficient*, MCC)⁶, apresentando os resultados de forma agregada. Em seguida, desagregamos as métricas de precisão e revocação por classe, a fim de identificar quais formas apresentaram maior dificuldade para os classificadores. Esta análise também nos permite observar o impacto de fatores como a separabilidade linear entre classes e o desbalanceamento de amostras, especialmente no caso da classe “Outra”.

No segundo grupo de experimentos, individualizamos a análise dos classificadores a partir das matrizes de confusão em que cada modelo alcançou sua maior acurácia. Essa abordagem possibilita examinar com mais profundidade como os erros se distribuem entre as classes, destacando padrões recorrentes de confusão e particularidades no comportamento de cada classificador. Além disso, a análise dessas matrizes nos permite observar o efeito da complexidade da superfície de decisão (linear ou não linear) sobre a capacidade de generalização dos modelos.

Por fim, no terceiro grupo de experimentos, comparamos a metodologia proposta com abordagens tradicionais da literatura, comumente utilizadas para descrever formas em imagens binárias. Consideramos, neste contexto, os Momentos Estatísticos, Momentos Centrais, Momentos de Hu, LBP e GLCM. Conduzimos a comparação por meio do teste estatístico de Wilcoxon ([WILCOXON, 1945](#)), considerando as métricas de acurácia e MCC,

⁶ O Apêndice C descreve as métricas de avaliação utilizadas nesta dissertação.

com o objetivo de determinar se os atributos derivados das medidas geométricas simples são estatisticamente equivalentes ou superiores às abordagens tradicionais.

As análises presentes nesta seção fornecem subsídios importantes para compreender os pontos fortes e as limitações da proposta inicial, bem como direcionar aprimoramentos metodológicos a serem considerados nas próximas etapas desta dissertação.

4.2.1 Projeto dos classificadores

Um fato importante que se deve salientar sobre atributos é que eles, normalmente, são entidades dimensionais (COSTA; CESAR JUNIOR, 2009). Por exemplo, a área de um comprimido pode ser medida em centímetros quadrados ou polegadas quadradas. A variação de unidades entre os atributos afeta diretamente a disposição dos padrões no espaço de características e, conseqüentemente, influencia a aprendizagem dos classificadores. Para atenuar esse problema, é comum aplicar um procedimento de normalização, também chamado de padronização, em que os atributos de todos os padrões são ajustados para uma mesma escala. Existem diferentes métodos de normalização, tais como *min-max* e *z-score* (ou normalização de média zero) (HAN; PEI; KAMBER, 2011). Para os experimentos deste capítulo, escolhemos a normalização de média zero, aplicada a cada atributo do conjunto de dados. Seja v o vetor que contém os valores de um mesmo atributo para todas as amostras, e v_i o valor desse atributo na amostra i . O valor normalizado $\hat{v}_i$ é calculado utilizando a equação

$$\hat{v}_i = \frac{v_i - \mu_v}{\sigma_v} \quad (4.1)$$

em que μ_v e σ_v correspondem à média e ao desvio padrão do atributo v , respectivamente. Assim, os dados estarão dispostos sob uma média com valor 0 e desvio padrão com valor 1.

Durante a etapa de treinamento, algoritmos de aprendizado de máquina aprendem um conjunto de parâmetros a partir dos dados. No entanto, há uma classe de parâmetros que independe do método utilizado para treinar o algoritmo, denominados hiperparâmetros (BENGIO, 2012). Esses parâmetros normalmente expressam propriedades que não podem ser aprendidas diretamente dos dados, como a complexidade do modelo ou a velocidade em que o modelo deve aprender. Dos classificadores utilizados nesse capítulo, apenas os classificadores LDA, QDA e NB não possuem hiperparâmetros.

O classificador k -NN possui como hiperparâmetros o número de vizinhos mais próximos (k) e a métrica de distância utilizada para medir a similaridade entre as amostras. Já o classificador MLP possui uma quantidade maior de hiperparâmetros, a saber: número de camadas ocultas, número de neurônios ocultos, função de ativação e número máximo de épocas do treinamento. Quando se utiliza o algoritmo de treinamento gradiente descendente estocástico (*Stochastic Gradient Descent*, SGD), deve-se considerar a taxa de aprendizagem

e o tamanho do lote (*batch*) como hiperparâmetros do MLP. O classificador ELM é semelhante ao MLP, porém possui apenas dois hiperparâmetros: número de neurônios ocultos e função de ativação. No caso do classificador SVM, os hiperparâmetros são o parâmetro de regularização C e qualquer parâmetro relativo à função de *kernel* escolhida, como a abertura da gaussiana σ , para funções de *kernel* gaussiano, ou o grau do polinômio d , para funções de *kernel* polinomial.

A escolha correta dos hiperparâmetros de um classificador tem relação direta com seu desempenho. Todavia, como não conhecemos os valores ótimos desses hiperparâmetros, devemos empregar uma seleção de modelos que nos forneça o melhor valor para cada hiperparâmetro especificado anteriormente. Entre as diferentes técnicas para busca de hiperparâmetros propostas na literatura, uma das mais populares é a busca em grade combinada com validação cruzada. O Apêndice B descreve o funcionamento da busca em grade e das técnicas de validação cruzada mais comuns.

Em todos os experimentos deste capítulo, utilizamos uma combinação das técnicas de validação cruzada *holdout* e *k-fold*. Primeiramente, com a técnica *holdout*, dividimos o conjunto de dados de modo que 80% seja utilizado durante o treinamento dos classificadores, e 20% seja utilizado para estimar as métricas de desempenho. Durante a etapa de treinamento de cada classificador, aplicamos a técnica *k-fold*, com cinco divisões ($k = 5$) para definir os valores ótimos dos hiperparâmetros. Particularmente:

- a) para o classificador k -NN, variamos o número de vizinhos mais próximos k no intervalo $[1, 3]$ e fixamos a distância euclidiana como função de similaridade;
- b) para o classificador MLP, definimos o número de camadas ocultas em 1, variando o número de neurônios ocultos no intervalo $[10, 15]$. Para todo neurônio oculto, a função de ativação é a tangente hiperbólica (*tanh*). Treinamos o classificador com o algoritmo SGD utilizando um lote de tamanho 200 e taxa de aprendizagem fixa de $1 \cdot 10^{-3}$. Por fim, fixamos o número máximo de épocas em 300;
- c) para o classificador ELM, definimos o número de camadas ocultas em 1, variando o número de neurônios ocultos no intervalo $[10, 15]$. Para todo neurônio oculto, a função de ativação é a tangente hiperbólica (*tanh*). Treinamos o classificador ELM utilizando sua formulação padrão;
- d) para o classificador SVM, utilizamos as funções de *kernel* linear (SVM/LIN), gaussiana (SVM/RBF) e polinomial (SVM/POL). Em particular, variamos os valores da constante de regularização C no intervalo $[2^1, 2^6]$, a abertura da gaussiana σ no intervalo $[2^{-5}, 2^6]$ e o grau do polinômio no intervalo $[2, 5]$, respectivamente. Treinamos cada classificador SVM com o algoritmo *Sequential Minimal Optimization* (SMO) (PLATT, 1999) e abordagem um-contra-um (KNERR; PERSONNAZ; DREYFUS, 1990).

A Tabela 4 apresenta um resumo dos hiperparâmetros de cada classificador que utilizamos nos experimentos, exceto para os classificadores LDA, QDA e NB.

Tabela 4 – Hiperparâmetros dos classificadores k -NN, MLP, ELM, SVM/LIN, SVM/POL e SVM/RBF.

Classificador	Parâmetro	Valores
k -NN	Número de vizinhos (k)	[1, 3]
	Função de similaridade	Distância euclidiana
MLP	Função de ativação	$\tanh$
	Limiar de tolerância para otimização	$1 \cdot 10^{-6}$
	Número de amostras por lote	200
	Número de neurônios ocultos	[10, 15]
	Número máximo de épocas	300
	Taxa de aprendizagem	$1 \cdot 10^{-3}$
ELM	Função de ativação	$\tanh$
	Número de neurônios ocultos	[10, 15]
SVM/LIN	Parâmetro de regularização (C)	$[2^1, 2^6]$
SVM/POL	Grau do polinômio (d)	[2, 5]
SVM/RBF	Abertura da função gaussiana (σ)	$[2^{-5}, 2^6]$

Fonte: Autor.

Os experimentos consistem em 30 simulações realizadas a partir da seleção aleatória dos subconjuntos de treinamento e teste. Para garantir uma análise justa de desempenho, treinamos os classificadores nas mesmas 30 divisões aleatórias do conjunto de dados. A seguir, apresentamos a análise dos resultados desses experimentos, considerando a média e o desvio padrão das 30 simulações.

4.2.2 Análise geral dos resultados

A Tabela 5 sumariza os desempenhos dos classificadores na identificação de formas dos comprimidos. No contexto de implementação, acurácia, precisão e revocação variam no intervalo numérico $[0, 1]$. No entanto, optamos por representar essas métricas em escala de porcentagem, visto que possui um apelo maior ao leitor. Já o coeficiente de correlação de Matthews, varia no intervalo numérico $[-1, 1]$.

Tabela 5 – Resultados dos classificadores no problema de identificação de formas, para as métricas: acurácia, MCC, precisão e revocação. Os valores de cada métrica são descritos em termos da média e desvio padrão, após 30 realizações de treinamento e teste. Destacamos os melhores desempenhos com fonte em negrito e cor de fundo laranja.

Classificador	Acurácia (%)	MCC	Precisão (%)	Revocação (%)
LDA	91,49 ± 1,32	0,87 ± 0,02	90,51 ± 2,19	91,49 ± 1,32
QDA	90,11 ± 1,53	0,86 ± 0,02	94,68 ± 1,16	90,11 ± 1,53
NB	63,37 ± 2,15	0,44 ± 0,03	60,95 ± 2,53	63,38 ± 2,16
k -NN	92,83 ± 1,24	0,89 ± 0,02	92,84 ± 1,24	92,83 ± 1,24
MLP	87,47 ± 2,14	0,81 ± 0,03	84,38 ± 2,33	87,47 ± 2,14
ELM	84,88 ± 2,53	0,76 ± 0,04	81,79 ± 2,60	84,88 ± 2,53
SVM/LIN	95,71 ± 0,84	0,93 ± 0,01	95,46 ± 1,05	95,72 ± 0,84
SVM/POL	93,27 ± 1,15	0,90 ± 0,02	90,27 ± 1,81	93,27 ± 1,15
SVM/RBF	95,22 ± 1,13	0,93 ± 0,02	94,35 ± 2,29	95,22 ± 1,13

Fonte: Autor.

Analisando a acurácia média dos classificadores, notamos o melhor desempenho dos classificadores SVM. Em especial, o SVM/LIN obteve a maior acurácia média entre todos os classificadores. As redes neurais também alcançaram bons resultados, com destaque para o MLP, que acertou em média 87,47% dos padrões de teste. Outro ponto positivo da análise são os resultados dos classificadores estatísticos LDA, QDA e k -NN, com acertos médios superiores aos das redes neurais MLP e ELM. Já o classificador NB destoa consideravelmente em relação aos outros classificadores, com 63,37% de acurácia média. Esse resultado sugere, a princípio, que a suposição de independência entre os atributos de forma é inadequada.

Os desempenhos dos classificadores lineares SVM/LIN e LDA nos permitem, ainda, inferir que os atributos de forma criam um espaço em que os padrões são, aproximadamente, linearmente separáveis. No caso particular do classificador SVM/LIN, os padrões estão dispostos de modo que é possível construir um hiperplano separador ótimo capaz de discriminar corretamente, em média, 95,71% dos padrões não vistos durante o treinamento. No caso do LDA, é interessante destacar que o uso de uma matriz de covariância única para todas as classes parece não prejudicar substancialmente seu desempenho, posto que alcançou uma acurácia média de 91,49%.

A complexidade da superfície de decisão de um classificador é um outro importante objeto de análise. De modo geral, constatamos que não é necessário construir superfícies

complexas para alcançar elevadas taxas médias de acerto. Em particular, o classificador QDA produz superfícies quadráticas capazes de classificar corretamente, em média, 90,11% dos padrões de teste. Indo além, as acurácias médias dos classificadores LDA e QDA sugerem que a suposição de que as densidades condicionais das classes são distribuições gaussianas multivariadas é adequada para o problema.

Não obstante, há fortes indícios de que os atributos de forma fornecem informações suficientes para a estimação da estrutura de probabilidade do problema. Fazemos essa análise com base nas acurácias médias dos classificadores k -NN, MLP, ELM e SVM. Entre esses, a rede ELM tem o menor desempenho, classificando corretamente, em média, 84,88% dos padrões de teste. Por fim, apesar de não possuir o melhor desempenho entre os classificadores, o k -NN certamente merece destaque, pois obtém uma taxa média de 90,11% de acerto nos padrões de teste.

Os resultados do MCC apresentam um ponto de vista alternativo sobre o desempenho dos classificadores. Em geral, os classificadores alcançaram elevadas taxas de correlação positiva entre suas predições e os rótulos verdadeiros. Os classificadores SVM novamente merecem destaque por obterem valores acima, em média, de 0,9. Com desempenhos ligeiramente inferiores estão os classificadores k -NN, LDA e QDA. Já as redes MLP e ELM possuem desempenho intermediário em torno de 0,81 e 0,76, respectivamente. O classificador NB destoa dos demais, apresentando uma correlação positiva substancialmente menor, com média de 0,44. Contudo, considerando todos os resultados, observamos que os classificadores foram consistentes em relação à aleatorização dos conjuntos de treinamento e teste, e robustos ao desbalanceamento das classes.

Precisão e revocação são métricas de avaliação para classificadores binários. Isto posto, separamos a análise dos resultados de ambas as métricas em duas partes. Primeiramente, discutimos os valores apresentados na Tabela 5, que são médias ponderadas, por classe, dos valores das métricas.

Em princípio, precisão e revocação nos permitem analisar o desempenho de um classificador sob a perspectiva de relevância. Uma avaliação global da precisão dos classificadores nos indica uma baixa incidência de falsos positivos. Isso significa que há poucos casos em que comprimidos preditos com uma classe pertenciam a outra. No caso particular do classificador NB, a taxa de falsos positivos varia entre 36% e 42%. Quando realizamos a mesma avaliação global da revocação dos classificadores, observamos que há uma baixa ocorrência de falsos negativos. Isso significa que há poucos erros de identificação dentro de uma mesma classe. Contudo, o classificador NB novamente destoa dos outros classificadores, com taxa de falsos negativos variando entre 35% e 40%. O uso prático de precisão e revocação, especialmente em contextos médicos, varia bastante e, portanto, não há como aferir se o NB é um classificador contraindicado para o problema-alvo da nossa pesquisa.

Na Tabela 6, apresentamos uma visão mais completa sobre as métricas de precisão

e revocação. De modo geral, os classificadores atingem altas taxas de precisão e revocação para as classes “Cápsula”, “Oval” e “Circular”. Em particular, percebemos que as menores ocorrências de falsos positivos e falsos negativos ocorrem na classe “Circular”. Alta precisão e alta revocação juntas indicam que os classificadores estão, tanto identificando corretamente a maioria dos comprimidos com essas formas (alta revocação), quanto evitando erros ao identificar essas formas (alta precisão).

Tabela 6 – Resultados, por classe, das métricas precisão e revocação. Os valores de cada métrica são descritos em termos da média e desvio padrão, após 30 realizações de treinamento e teste. Destacamos os melhores desempenhos com fonte em negrito e cor de fundo laranja.

Métrica	Classificador	Classe			
		Cápsula	Oval	Circular	Outra
Precisão	LDA	90,11 ± 3,45	93,82 ± 1,84	90,48 ± 2,09	56,67 ± 49,55
	QDA	96,46 ± 2,35	96,52 ± 2,33	98,30 ± 0,85	25,06 ± 5,71
	NB	60,14 ± 5,21	55,33 ± 4,87	69,19 ± 2,22	13,27 ± 9,37
	k-NN	94,94 ± 2,31	93,28 ± 2,19	94,90 ± 1,62	52,79 ± 12,96
	MLP	90,02 ± 4,33	86,88 ± 4,89	87,25 ± 2,71	0,00 ± 0,00
	ELM	86,87 ± 5,16	81,91 ± 4,52	86,47 ± 3,94	0,00 ± 0,00
	SVM/LIN	94,44 ± 2,53	94,99 ± 1,68	97,10 ± 1,05	83,25 ± 16,43
	SVM/POL	95,85 ± 2,18	91,15 ± 2,16	94,40 ± 1,67	6,94 ± 14,76
	SVM/RBF	95,07 ± 2,74	94,93 ± 1,50	96,11 ± 1,89	64,96 ± 33,04
Revocação	LDA	99,22 ± 0,82	86,19 ± 2,94	99,72 ± 0,32	6,10 ± 6,73
	QDA	97,76 ± 2,02	76,46 ± 4,08	99,52 ± 0,54	66,15 ± 10,17
	NB	76,51 ± 5,69	29,35 ± 3,25	88,46 ± 2,07	8,94 ± 7,22
	k-NN	95,19 ± 3,01	91,59 ± 2,06	96,46 ± 1,55	51,05 ± 12,90
	MLP	82,11 ± 9,45	83,32 ± 3,97	100,00 ± 0,00	0,00 ± 0,00
	ELM	74,05 ± 8,87	81,49 ± 6,16	98,69 ± 1,76	0,00 ± 0,00
	SVM/LIN	98,55 ± 1,77	96,26 ± 1,61	99,96 ± 0,21	27,69 ± 8,30
	SVM/POL	92,58 ± 3,46	95,28 ± 1,91	99,71 ± 0,36	1,43 ± 2,96
	SVM/RBF	97,64 ± 2,06	96,14 ± 1,59	99,46 ± 0,82	27,76 ± 19,87

Fonte: Autor.

Um ponto que é necessário destacarmos, e que vale para a análise de ambas as métricas, é a identificação de comprimidos da classe “Outra”. Para os classificadores MLP

e ELM, tanto precisão quanto revocação revelam a dificuldade desses classificadores com padrões dessa classe. No caso particular da precisão, esse comportamento também pode ser observado nos classificadores QDA, NB e SVM/POL, todos com taxa de falsos positivos superior, em média, a 75%. Os únicos classificadores que apresentam desempenho no mínimo razoável são: LDA, k -NN, SVM/LIN e SVM/RBF. Em especial, o classificador SVM/LIN é o único a possuir taxa de falsos positivos inferior, em média, a 17%.

Uma avaliação semelhante pode ser feita com a revocação. Os classificadores QDA e k -NN são os únicos que alcançam desempenho razoável, ao passo que os outros demonstram extrema dificuldade com padrões da classe “Outra”. A baixa taxa de revocação nesse caso em particular, sugere que esses classificadores estão deixando de identificar corretamente muitas formas reais de comprimidos. Em especial, os classificadores MLP e ELM não conseguem identificar corretamente nenhum comprimido com formas da classe “Outra”. Uma justificativa para tal pode ser o baixo número de amostras que compõe essa classe.

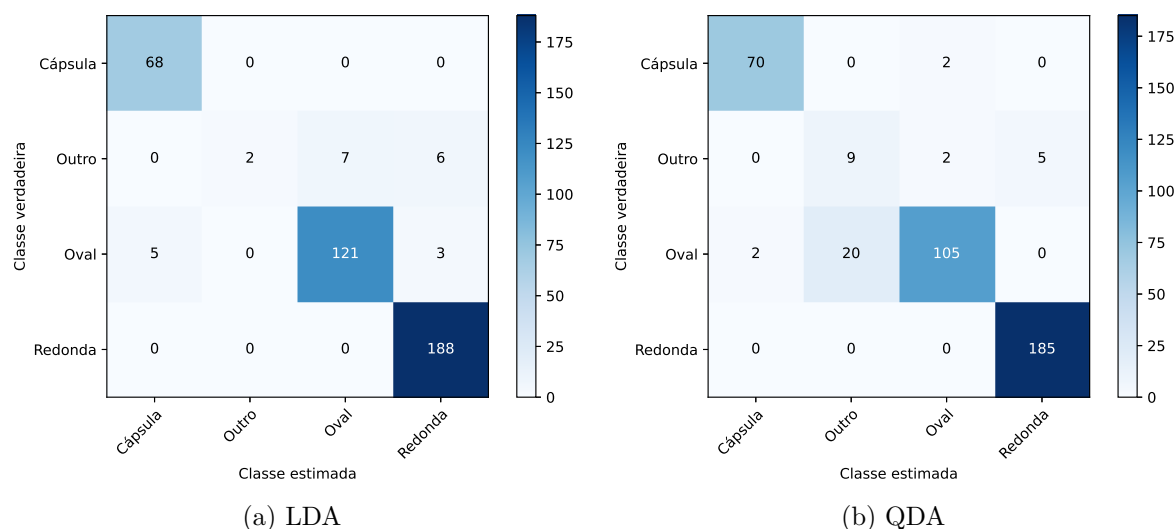
4.2.3 Análise individual dos classificadores

Para melhor entender o desempenho individual dos classificadores, exploramos a seguir suas matrizes de confusão em que esses alcançaram maior acurácia. A Figura 11a mostra a matriz de confusão do classificador LDA. Analisando a matriz, nota-se que não houve erros de identificação para as classes “Cápsula” e “Circular”. Um segundo ponto que destacamos é que, embora tenha acertado as classes da maioria dos padrões “ovais”, ainda houve dificuldade na discriminação de alguns comprimidos, ora classificando-os como “cápsulas”, ora como “circulares”. Por fim, apesar de ter alcançado bom desempenho médio em todas as métricas, o LDA errou as classes de todos os comprimidos da classe “Outra”.

A Figura 11b mostra a matriz de confusão do classificador QDA. Analisando a matriz, nota-se que o QDA acertou a identificação de todos os comprimidos da classe “Circular”. Diferente do LDA, o classificador QDA cometeu erros de identificação na classe “Cápsula”, discriminando 2 comprimidos como “ovais”. A dificuldade que identificamos no classificador LDA é mais evidente no QDA, que discriminou 20 comprimidos ovais como “outros”. Uma justificativa para essa maior dificuldade do QDA em relação ao LDA pode residir na superfície de decisão quadrática, que é mais complexa que a linear. Por fim, o QDA também não conseguiu acertar nenhum padrão da classe “Outra”.

A Figura 12a mostra a matriz de confusão do classificador NB. A análise da matriz reforça as observações anteriores sobre o desempenho desse classificador, que comete muito erros, independente da classe. A forte/ingênua suposição de independência entre os atributos aparenta ser, de fato, inadequada para o problema de identificação de formas. A classe “Oval” novamente é a que compreende o maior número de predições incorretas, com quase 70% das amostras classificadas incorretamente. Ademais, o classificador atribui erroneamente 13 comprimidos da classe “Cápsula” a outras formas, totalizando uma taxa

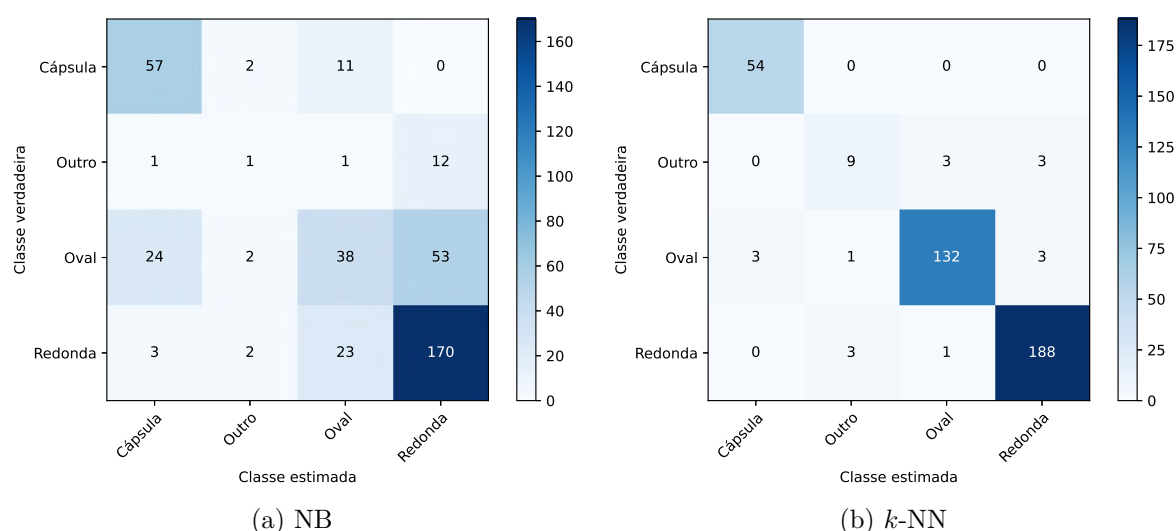
Figura 11 – Matrizes de confusão em que os classificadores LDA e QDA obtiveram maior acurácia no conjunto de teste.



Fonte: Autor.

de falsos negativos próxima a 18%. Surpreendentemente, o NB ainda conseguiu classificar corretamente 1 comprimido da classe “Outra”.

Figura 12 – Matrizes de confusão em que os classificadores NB e k -NN obtiveram maior acurácia no conjunto de teste.



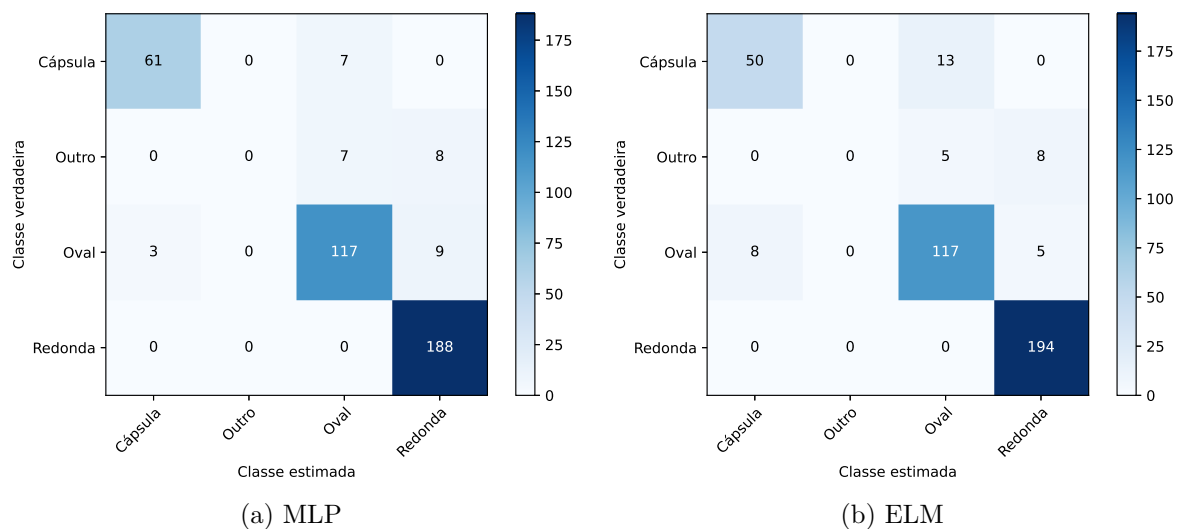
Fonte: Autor.

A Figura 12b mostra a matriz de confusão do classificador k -NN. Analisando a matriz, nota-se que não houve erros de identificação para comprimidos da classe “Cápsula”. A taxa de falsos negativos para a classe “Oval” é menor que a obtida pelo classificador LDA. No entanto, o classificador k -NN classifica incorretamente 4 comprimidos da classe “Circular”. Enfim, o k -NN também não conseguiu acertar nenhum padrão da classe “Outra”.

A Figura 13a mostra a matriz de confusão do classificador MLP. Analisando a matriz, nota-se que não houve erros de identificação para a classe “Circular”. Ademais, a taxa de falsos negativos para a classe “Cápsula” é inferior à dos classificadores estatísticos (com exceção do NB). Assim como o classificador LDA, o MLP também encontra um pouco de dificuldade ao classificar comprimidos da classe “Oval”, ora classificando-os como “cápsulas”, ora como “circulares”. Por fim, o MLP também não conseguiu acertar nenhum padrão da classe “Outra”.

A Figura 13b mostra a matriz de confusão do classificador ELM. Assim como os classificadores LDA, QDA e MLP, o classificador ELM não cometeu erros para a classe “Circular”. Outrossim, a taxa de falsos negativos para a classe “Cápsula” é inferior à dos classificadores estatísticos (com exceção do NB) e MLP. A ELM também encontra um pouco de dificuldade ao classificar comprimidos da classe “Oval”, ora discriminando-os como “cápsulas”, ora como “circulares”. Por fim, a ELM também não conseguiu acertar nenhum padrão da classe “Outra”.

Figura 13 – Matrizes de confusão em que os classificadores MLP e ELM obtiveram maior acurácia no conjunto de teste.



Fonte: Autor.

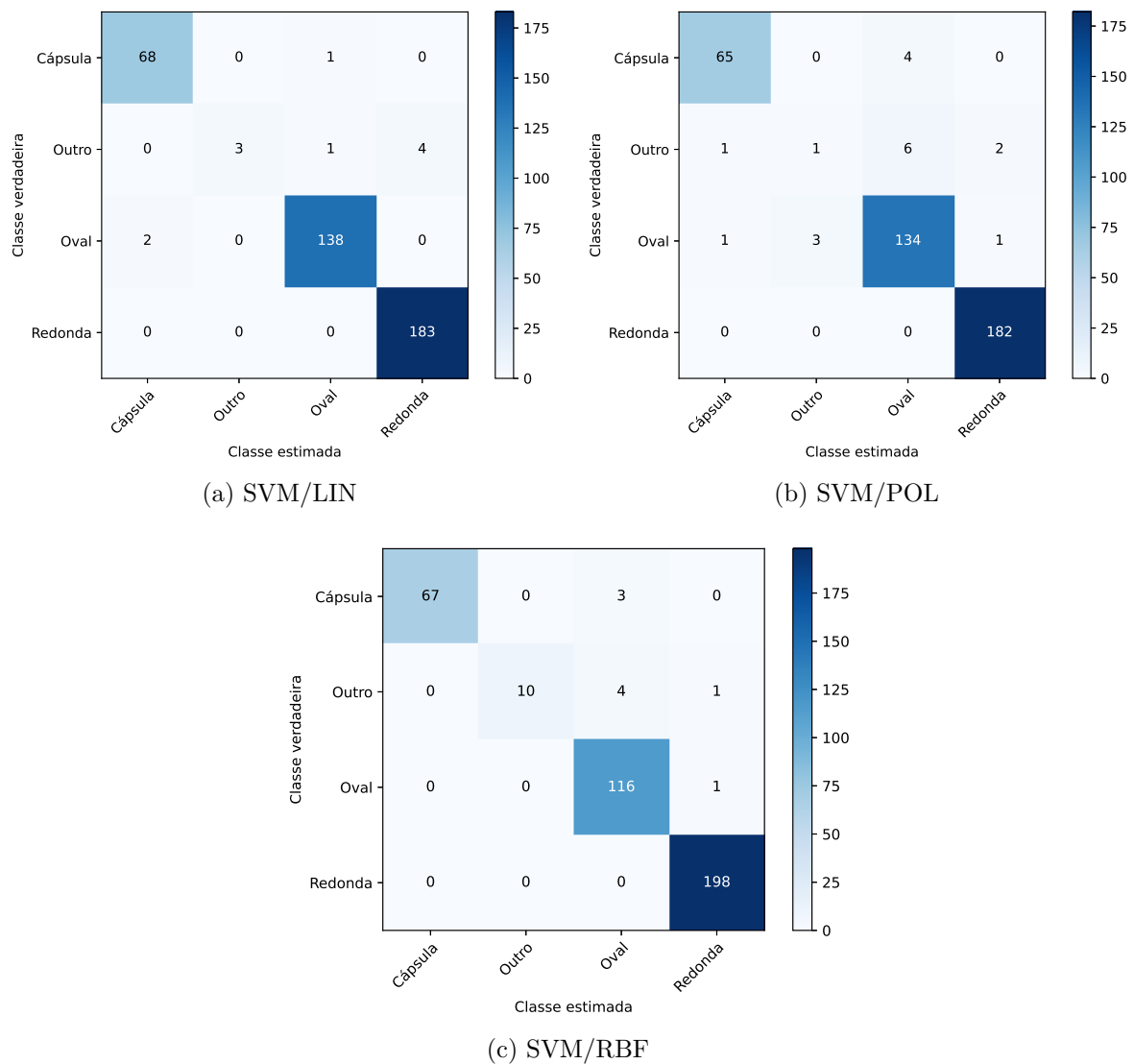
A Figura 14a mostra a matriz de confusão do classificador SVM/LIN. Analisando a matriz, nota-se que não houve erros de identificação para a classe “Circular”. Ademais, o classificador cometeu apenas 1 erro para a classe “Cápsula”, e 2 erros para a classe “Oval”. Por fim, a SVM/LIN também não conseguiu acertar nenhum padrão da classe “Outra”.

A Figura 14b mostra a matriz de confusão do classificador SVM/POL. Semelhante ao SVM/LIN, a SVM/POL não cometeu erros de identificação para a classe “Circular”. No entanto, o classificador cometeu mais erros para a classe “Oval” que a SVM/LIN. Outrossim, a SVM/POL cometeu 4 erros para a classe “Cápsula”. Por fim, a SVM/POL

também não conseguiu acertar nenhum padrão da classe “Outra”.

A Figura 14c mostra a matriz de confusão do classificador SVM/RBF. Analisando a matriz, nota-se que não houve erros de identificação para a classe “Circular”. Ademais, a SVM/RBF cometeu menos erros para a classe “Oval” que as SVM/LIN e SVM/POL. Para a classe “Cápsula”, a SVM/RBF cometeu 4 erros, discriminando todos os padrões como “ovais”. Por fim, a SVM/RBF também não conseguiu acertar nenhum padrão da classe “Outra”.

Figura 14 – Matrizes de confusão em que os classificadores SVM/LIN, SVM/POL e SVM/RBF obtiveram maior acurácia no conjunto de teste.



Fonte: Autor.

4.2.4 Comparação com outros extratores de atributos

Nesta seção, realizamos uma análise comparativa da nossa metodologia com alguns extratores clássicos da literatura, a saber: Momentos Estatísticos, Momentos Centrais, Momentos de Hu e LBP e GLCM. Para garantir a robustez da análise comparativa, aplicamos o teste estatístico de Wilcoxon pareado sobre as acurácias e os MCCs obtidos após 30 simulações independentes para cada classificador. Os resultados estão sumarizados nas Tabelas 7 e 8.

Especificamente para o extrator LBP, utilizamos 24 vizinhos igualmente espaçados em um raio de oito pixels e a versão “uniforme” sobre as imagens em escala de cinza. Essa configuração foi escolhida por oferecer um bom equilíbrio entre discriminação de padrões de textura e custo computacional. O histograma resultante foi normalizado e utilizado como vetor de atributos.

De forma geral, os atributos geométricos propostos apresentaram desempenho superior ou estatisticamente equivalente aos métodos comparados, com destaque para os classificadores SVM/LIN e SVM/RBF, que mantiveram altos valores de acurácia (95,71% e 95,22%, respectivamente) e MCC (0,93 em ambos os casos).

O desempenho inferior dos Momentos de Hu e GLCM pode ser atribuído à sua maior sensibilidade a ruídos de contorno e variações no alinhamento das imagens segmentadas. O LBP, embora mais robusto em problemas de textura, apresentou desempenho instável, o que reforça a inadequação de descritores locais para um problema centrado na morfologia global dos objetos.

A análise dos Momentos Centrais e Estatísticos revela desempenho competitivo em alguns classificadores, mas ainda assim inferior ao dos atributos propostos, principalmente em redes neurais. Essa observação sugere que os atributos geométricos oferecem uma representação mais compacta e semanticamente alinhada com o domínio do problema, além de serem computacionalmente mais simples de extrair.

Vale destacar que os métodos clássicos testados possuem dimensionalidade significativamente maior (GLCM, por exemplo, com dezenas de variáveis), o que impacta negativamente o tempo de treinamento e pode aumentar a propensão ao sobreajuste em contextos com conjuntos de dados pequenos ou desbalanceados, como a classe “Outra”.

Esses resultados indicam que a estratégia de seleção de atributos baseada em conhecimento do domínio foi eficaz, e que descritores geométricos simples e interpretáveis são mais apropriados para tarefas de identificação de formas de comprimidos, superando técnicas genéricas de análise de imagem em desempenho, eficiência computacional e interpretabilidade.

Tabela 7 – Resultados do teste estatístico de Wilcoxon, com base nos valores de acurácia (%), para aferição de equivalência ou não equivalência entre o método que propomos e os extratores: Momentos Estatísticos (Mom. Es.), Momentos Centrais (Mom. Ce.), Momentos de Hu (Mom. Hu), LBP e GLCM. Os símbolos ✓ e ✗ indicam equivalência ou não equivalência, respectivamente.

Classificador	Extrator					
	Nosso	Mom. Es.	Mom. Ce.	Mom. Hu	LBP	GLCM
LDA	91,49 ± 1,32	86,16 ± 1,66 ✗	72,97 ± 1,99 ✗	72,84 ± 1,95 ✗	80,66 ± 1,86 ✗	67,87 ± 2,17 ✗
QDA	90,11 ± 1,53	97,21 ± 0,76 ✗	82,47 ± 1,92 ✗	73,25 ± 1,84 ✗	78,88 ± 2,11 ✗	68,06 ± 2,07 ✗
NB	63,37 ± 2,15	52,42 ± 2,08 ✗	73,32 ± 1,65 ✗	54,57 ± 2,46 ✗	69,07 ± 1,73 ✗	58,08 ± 1,95 ✗
k-NN	92,83 ± 1,24	96,28 ± 0,88 ✗	85,60 ± 1,77 ✗	97,58 ± 0,72 ✗	79,53 ± 1,95 ✗	72,57 ± 2,26 ✗
MLP	87,47 ± 2,14	67,65 ± 2,87 ✗	75,28 ± 2,05 ✗	74,90 ± 2,37 ✗	79,56 ± 1,84 ✗	67,25 ± 1,92 ✗
ELM	84,88 ± 2,53	94,61 ± 1,23 ✗	80,81 ± 1,85 ✗	87,33 ± 4,16 ✗	77,08 ± 1,83 ✗	69,16 ± 2,15 ✗
SVM/LIN	95,71 ± 0,84	93,92 ± 1,12 ✗	93,97 ± 1,08 ✗	89,08 ± 2,01 ✗	83,59 ± 2,03 ✗	71,12 ± 2,09 ✗
SVM/POL	93,27 ± 1,15	69,23 ± 2,43 ✗	77,67 ± 1,60 ✗	70,30 ± 2,25 ✗	80,20 ± 1,91 ✗	66,01 ± 2,30 ✗
SVM/RBF	95,22 ± 1,13	96,78 ± 0,86 ✗	95,56 ± 1,01 ✓	96,92 ± 0,88 ✗	83,49 ± 1,47 ✗	74,24 ± 2,34 ✗

Fonte: Autor.

Tabela 8 – Resultados do teste estatístico de Wilcoxon, com base nos valores do MCC, para aferição de equivalência ou não equivalência entre o método que propomos e os extratores: Momentos Estatísticos (Mom. Es.), Momentos Centrais (Mom. Ce.), Momentos de Hu (Mom. Hu), LBP e GLCM. Os símbolos ✓ e ✗ indicam equivalência ou não equivalência, respectivamente.

Classificador	Extrator					
	Nosso	Mom. Es.	Mom. Ce.	Mom. Hu	LBP	GLCM
LDA	0,87 ± 0,02	0,79 ± 0,02 ✗	0,61 ± 0,03 ✗	0,27 ± 0,05 ✗	0,71 ± 0,03 ✗	0,49 ± 0,04 ✗
QDA	0,86 ± 0,02	0,96 ± 0,01 ✗	0,75 ± 0,03 ✗	0,58 ± 0,03 ✗	0,68 ± 0,03 ✗	0,53 ± 0,03 ✗
NB	0,44 ± 0,03	0,29 ± 0,03 ✗	0,60 ± 0,02 ✗	0,63 ± 0,03 ✗	0,55 ± 0,02 ✗	0,38 ± 0,03 ✗
k-NN	0,89 ± 0,02	0,94 ± 0,01 ✗	0,78 ± 0,03 ✗	0,96 ± 0,01 ✗	0,68 ± 0,03 ✗	0,58 ± 0,04 ✗
MLP	0,81 ± 0,03	0,49 ± 0,05 ✗	0,62 ± 0,03 ✗	0,62 ± 0,04 ✗	0,68 ± 0,03 ✗	0,49 ± 0,03 ✗
ELM	0,76 ± 0,04	0,92 ± 0,02 ✗	0,72 ± 0,02 ✗	0,81 ± 0,06 ✗	0,65 ± 0,03 ✗	0,52 ± 0,03 ✗
SVM/LIN	0,93 ± 0,01	0,91 ± 0,02 ✗	0,91 ± 0,02 ✗	0,83 ± 0,03 ✗	0,75 ± 0,03 ✗	0,55 ± 0,03 ✗
SVM/POL	0,90 ± 0,02	0,53 ± 0,04 ✗	0,68 ± 0,02 ✗	0,57 ± 0,03 ✗	0,70 ± 0,03 ✗	0,49 ± 0,04 ✗
SVM/RBF	0,93 ± 0,02	0,95 ± 0,01 ✗	0,93 ± 0,02 ✓	0,95 ± 0,01 ✗	0,74 ± 0,02 ✗	0,60 ± 0,04 ✗

Fonte: Autor.

4.3 Considerações finais

Após analisar os desempenhos dos classificadores na discriminação das formas dos comprimidos, destacamos a seguir nossas principais considerações:

- a) os modelos lineares alcançaram desempenho superior em relação aos demais, sugerindo que o problema apresenta separabilidade aproximadamente linear;
- b) de modo geral, a proposta apresentou melhor desempenho aliado a menor complexidade quando comparada a outros métodos de extração de atributos, o que evidencia sua eficiência e aplicabilidade prática;
- c) apesar de o baixo número de padrões da classe “Outra” não ter comprometido significativamente o desempenho de todos os classificadores, esse desbalanceamento impactou a capacidade de generalização, sobretudo dos classificadores MLP e ELM;
- d) a suposição de independência entre os atributos de forma mostrou-se inadequada, indicando que há correlações relevantes entre as variáveis que precisam ser consideradas nos modelos probabilísticos;
- e) a distribuição Gaussiana multivariada mostrou-se apropriada porque capturou correlações entre os atributos e apresentou boa aderência aos dados observados. Além disso, o desempenho obtido pelos classificadores baseados nessa modelagem reforça que a estrutura de probabilidade assumida é compatível com a natureza do problema;
- f) os atributos de forma fornecem informações suficientes para a estimação de uma estrutura de probabilidade coerente com o problema de classificação, permitindo discriminar adequadamente as diferentes formas dos comprimidos.

5 Identificação de comprimidos com poucos exemplos e autossupervisão

A identificação de comprimidos a partir de imagens é uma tarefa de visão computacional de granularidade fina que apresenta desafios singulares. Além da semelhança visual entre diferentes comprimidos, muitas vezes com variações mínimas de forma, textura ou cores, há também uma limitação prática quanto à disponibilidade de conjuntos de dados rotulados de alta qualidade, especialmente no domínio biomédico. Esses obstáculos tornam métodos supervisionados tradicionais menos eficazes ou inviáveis em contextos reais, como a identificação automatizada de comprimidos em ambientes clínicos ou domiciliares.

Neste capítulo, propomos e avaliamos uma metodologia denominada SSL4PILL, que combina aprendizado autossupervisionado e aprendizado com poucos exemplos para mitigar essas dificuldades. A proposta parte do princípio de que representações visuais de alta qualidade, quando aprendidas previamente sem necessidade de rótulos, podem ser adaptadas com eficiência a tarefas específicas mesmo com um número mínimo de exemplos rotulados. Com isso, almejamos reduzir os custos de rotulação e, simultaneamente, aumentar a capacidade do sistema de generalizar para novos comprimidos que não estavam presentes durante o treinamento.

Além da descrição detalhada da metodologia, este capítulo também apresenta a configuração dos experimentos, incluindo os conjuntos de dados que utilizamos, os protocolos de avaliação que adotamos e os principais parâmetros de treinamento. Por fim, discutimos os resultados obtidos, comparando a eficácia do modelo proposto em relação a abordagens existentes da literatura, com destaque para o impacto do pré-treinamento autossupervisionado no desempenho final.

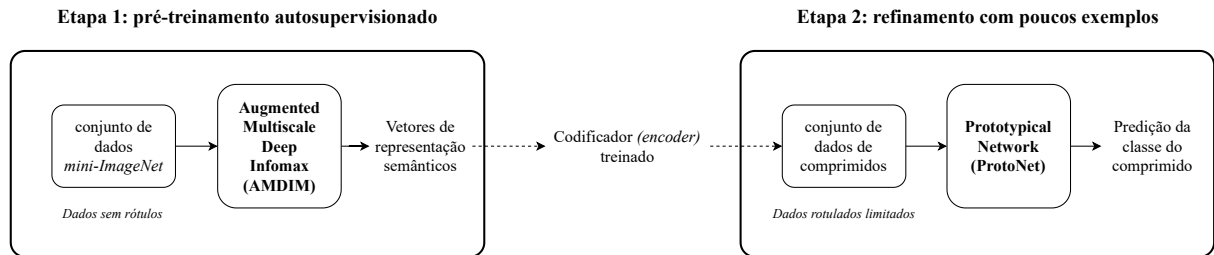
Os principais objetivos deste capítulo são:

- a) apresentar a metodologia SSL4PILL e justificar o uso conjunto de aprendizado autossupervisionado e aprendizado com poucos exemplos;
- b) demonstrar que o pré-treinamento autossupervisionado permite a extração de representações transferíveis para o domínio de imagens de comprimidos;
- c) avaliar empiricamente a eficácia da proposta em dois conjuntos de dados desafiadores e amplamente utilizados na área;
- d) e analisar os limites e as possibilidades de generalização da abordagem em tarefas de identificação de objetos biomédicos com baixa disponibilidade de rótulos.

5.1 SSL4PILL

Nós estruturamos a SSL4PILL em dois estágios principais. O primeiro utiliza o método *Augmented Multiscale Deep InfoMax* (AMDIM) (BACHMAN; HJELM; BUCHWALTER, 2019) para o pré-treinamento de uma rede codificadora (*encoder*) sobre um grande conjunto de imagens genéricas, explorando o aprendizado autossupervisionado baseado em contrastes multiescalares. O segundo estágio emprega o modelo *Prototypical Networks* (ProtoNet) (SNELL; SWERSKY; ZEMEL, 2017) para adaptar a rede codificadora à tarefa-alvo com poucos exemplos rotulados, por meio de uma estratégia de aprendizado métrico que simula classificações de baixa supervisão em múltiplos episódios. A Figura 15 ilustra a estrutura geral do SSL4PILL.

Figura 15 – Estrutura da metodologia SSL4PILL para identificação de comprimidos com poucos exemplos e autossupervisão.



Fonte: Autor.

A nossa proposta se apoia em uma estratégia de transferência de aprendizado entre os dois estágios. Inicialmente, treinamos a rede codificadora com AMDIM no *mini-ImageNet* (VINYALS *et al.*, 2016) como ponto de partida para o ProtoNet, que é refinado com imagens reais de comprimidos obtidas em condições não controladas.

Essa integração permite aproveitar representações genéricas de forma eficiente, reduzindo o risco de sobreajuste no domínio-alvo, especialmente relevante quando o número de exemplos rotulados é limitado. Ao unir as vantagens do aprendizado autossupervisionado com o aprendizado métrico, a metodologia SSL4PILL oferece uma solução robusta para a identificação de comprimidos em cenários com baixa disponibilidade de dados rotulados.

5.1.1 Estágio 1: pré-treinamento com autossupervisão

A etapa de aprendizado autossupervisionado baseia-se no modelo AMDIM, que estende a abordagem *Deep InfoMax* local (HJELM *et al.*, 2019). O AMDIM calcula a informação mútua entre múltiplas representações de características derivadas de diferentes visões aumentadas da mesma imagem. O princípio fundamental desse método é que a informação mútua entre características correspondentes de múltiplas visões captura

o conteúdo semântico essencial, mantendo invariância em relação às transformações superficiais (BACHMAN; HJELM; BUCHWALTER, 2019).

Cada imagem de entrada $\mathbf{x}$ é transformada em duas visões aumentadas $\mathbf{x}_a = t_a(\mathbf{x})$ e $\mathbf{x}_b = t_b(\mathbf{x})$, em que $t_a, t_b \sim \mathcal{T}$ representam transformações amostradas de uma distribuição de aumentos $\mathcal{T}$. Essa distribuição inclui transformações geométricas, como recorte aleatório e redimensionamento, e modificações de aparência, como variação de cor e conversão para tons de cinza. A Figura 16a ilustra esse processo de aumento de dados.

Para cada visão aumentada, uma rede codificadora f^7 extrai uma sequência hierárquica de mapas de características:

$$\{f_1(\mathbf{x}), f_2(\mathbf{x}), \dots, f_D(\mathbf{x})\},$$

em que cada $f_d(\mathbf{x}) \in \mathbb{R}^{H_d \times W_d \times C_d}$ representa as ativações de características na profundidade d , com dimensões espaciais $H_d \times W_d$ e C_d canais, e D representa o número total de profundidades consideradas no codificador.

Em seguida, o modelo amostra pares de características locais e globais das visões aumentadas e os avalia com uma função de perda contrastiva baseada na estimativa de contraste de ruído (*noise contrastive estimation*, NCE). Então, o modelo maximiza sistematicamente a informação mútua entre representações de características correspondentes nas visões aumentadas, enquanto minimiza a informação com características não correspondentes. O estimador padrão InfoNCE (OORD; LI; VINYALS, 2018) é dado por:

$$\mathcal{L}_{\text{InfoNCE}}(f_a, f_b) = -\log \frac{\exp(f_a \cdot f_b / \tau)}{\exp(f_a \cdot f_b / \tau) + \sum_{n=1}^N \exp(f_a \cdot f_n / \tau)}, \quad (5.1)$$

em que $f_a = f_d(\mathbf{x}_a)^{(i,j)}$ é a característica na posição (i, j) da visão aumentada $\mathbf{x}_a$, $f_b = f_d(\mathbf{x}_b)^{(i',j')}$ é a característica correspondente na visão aumentada $\mathbf{x}_b$, $\{f_n\}_{n=1}^N$ são amostras negativas do lote, e τ é o hiperparâmetro de temperatura⁸.

Essa estratégia multiescala desempenha um papel central na eficácia do AMDIM. Em vez de depender de um único nível de abstração, o modelo extrai mapas de características $\{f_d(\cdot)\}_{d=1}^D$ de múltiplas profundidades intermediárias do codificador e calcula informações mútuas entre eles. Para cada par de visões aumentadas $\mathbf{x}_a$ e $\mathbf{x}_b$, o AMDIM amostra características na mesma profundidade $d = d'$ (intra-escala) e entre profundidades diferentes $d \neq d'$ (entre-escalas), produzindo um conjunto abrangente de pares positivos:

$$\mathcal{P} = \left\{ \left(f_d(\mathbf{x}_a)^{(i,j)}, f_{d'}(\mathbf{x}_b)^{(i',j')} \right) : d, d' \in \{1, \dots, D\} \right\}. \quad (5.2)$$

⁷ O codificador f é uma rede neural profunda com parâmetros θ . Para simplificar a notação matemática, usamos f em vez de f_θ .

⁸ O hiperparâmetro de temperatura controla o quão “rigoroso” o modelo é em distinguir pares positivos (mesma imagem) de pares negativos (imagens diferentes).

O modelo atribui uma pontuação a cada par usando uma função de similaridade de produto interno, que estima a informação mútua entre as representações das diferentes visões aumentadas. O treinamento maximiza o objetivo esperado do InfoNCE neste conjunto. Então, a função de custo contrastiva multiescala completa se torna:

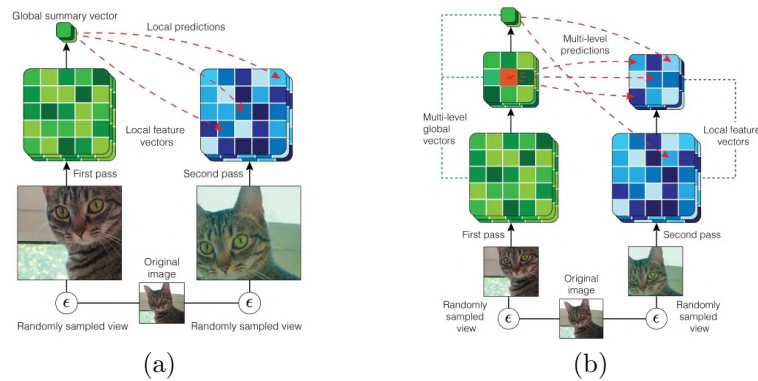
$$\mathcal{L}_{ssl} = \sum_{(f_a, f_b) \in \mathcal{P}} \mathcal{L}_{\text{InfoNCE}}(f_a, f_b). \quad (5.3)$$

Especificamente, o AMDIM utiliza três funções de custo, que se dividem em dois grupos:

- a) custos locais $\rightarrow$ custos locais, que maximizam informações mútuas entre características espacialmente correspondentes:
 - $\mathcal{L}_{\text{InfoNCE}}(f_5, f_5)$;
- b) custos globais $\rightarrow$ custos locais, que alinham características locais com representações de contexto global:
 - $\mathcal{L}_{\text{InfoNCE}}(f_1, f_5)$;
 - $\mathcal{L}_{\text{InfoNCE}}(f_1, f_7)$.

A formulação multiescala na Eq. 5.3 impõe explicitamente a consistência de características em todas as dimensões espaciais e níveis de abstração, estabelecendo uma hierarquia representacional abrangente. Fundamentalmente, ao otimizar múltiplas escalas conjuntamente, o AMDIM aprende representações que capturam simultaneamente características visuais de baixo nível (texturas, bordas) e conceitos semânticos de alto nível (partes de objetos, categorias). Essa estrutura estabelece uma base para tarefas posteriores, como a identificação de comprimidos com poucos exemplos, como propomos neste capítulo. A Figura 16b ilustra o modelo AMDIM completo.

Figura 16 – (a) Processo de aumento de dados que gera duas visualizações aumentadas $\mathbf{x}_a$ e $\mathbf{x}_b$ a partir de uma imagem original $\mathbf{x}$. (b) Arquitetura do AMDIM, com função de custo multiescala em visões geradas por aumento de dados.



Fonte: [Bachman, Hjelm e Buchwalter \(2019\)](#).

5.1.2 Estágio 2: refinamento com aprendizado de poucos exemplos

Após o pré-treinamento autossupervisionado do codificador f , o ajustamos à tarefa de identificação de comprimidos com o método de aprendizado métrico ProtoNet. O ProtoNet utiliza treinamento episódico, uma estratégia que simula a configuração de poucas tentativas no tempo de teste durante a etapa de treinamento.

Em cada episódio $\mathcal{E}_i$, define-se uma tarefa de classificação composta por dois subconjuntos:

- a) **conjunto de suporte** $\mathcal{S} = \{(\mathbf{x}_i, y_i)\}_{i=1}^{N \times K}$, contendo N amostras rotuladas de cada uma das K classes;
- b) **conjunto de consulta** $\mathcal{Q} = \{(\mathbf{x}_j, y_j)\}_{j=1}^{M \times K}$, contendo M amostras não rotuladas das mesmas K classes.

Para cada classe $k \in \{1, \dots, K\}$, o modelo calcula um vetor de protótipo $\mathbf{p}_k \in \mathbb{R}^n$ através da média dos vetores de representação de seus exemplos de suporte:

$$\mathbf{p}_k = \frac{1}{K} \sum_{(x_i, y_i) \in \mathcal{S}_k} f(x_i), \quad (5.4)$$

em que $\mathcal{S}_k \subset \mathcal{S}$ inclui todos os exemplos de suporte rotulados como classe k , e $f(x_i)$ denota a representação codificada da imagem x_i .

Para classificar uma amostra de consulta $x_q \in \mathcal{Q}$, o modelo compara seu vetor de representação com o vetor de protótipo de cada classe usando uma função de similaridade. Seguindo a formulação original da ProtoNet, adotamos o quadrado da distância euclidiana como função de similaridade:

$$d(f(x_q), \mathbf{p}_k) = \|f(x_q) - \mathbf{p}_k\|_2^2. \quad (5.5)$$

Então, o modelo computa o rótulo da imagem de consulta calculando uma distribuição de probabilidade sobre todos os protótipos usando a função *softmax*:

$$P(y = c \mid x_q) = \frac{\exp(-d(f(x_q), \mathbf{p}_c))}{\sum_{c'=1}^K \exp(-d(f(x_q), \mathbf{p}_{c'}))}. \quad (5.6)$$

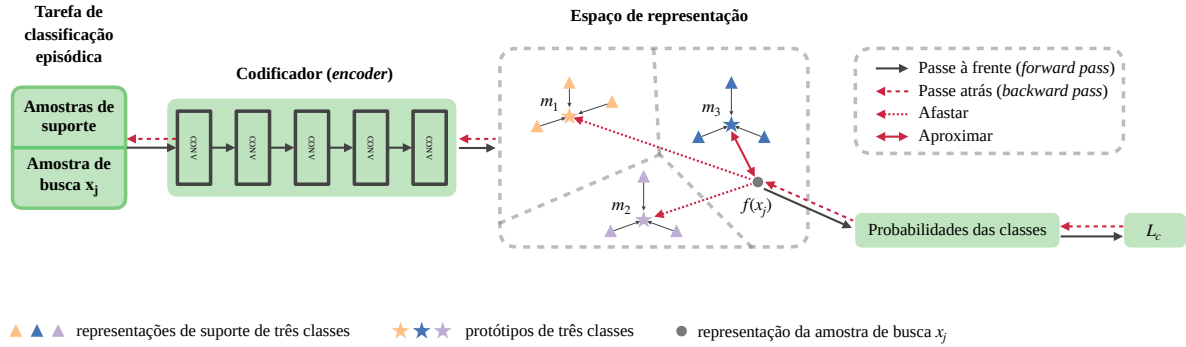
Esta formulação probabilística incentiva o modelo a atribuir maior confiança aos protótipos mais próximos no espaço de características.

Durante o treinamento, atualizamos os parâmetros do codificador minimizando a log-verossimilhança negativa sobre os rótulos verdadeiros de todos os exemplos de consulta:

$$\mathcal{L}_{\text{proto}} = - \sum_{(x_q, y_q) \in \mathcal{Q}} \log P(y = y_q \mid x_q). \quad (5.7)$$

Ao otimizar repetidamente essa função de custo ao longo de vários episódios, o ProtoNet aprende um espaço de características onde exemplos da mesma classe se agrupam em torno de um protótipo comum, permitindo uma classificação eficiente com supervisão mínima. A Figura 17 ilustra o processo de refinamento com ProtoNet.

Figura 17 – Etapa de refinamento da metodologia SSL4PILL com ProtoNet.



Fonte: Autor.

5.2 Configuração experimental

Nesta seção, descrevemos a configuração experimental que adotamos para avaliar a eficácia do modelo SSL4PILL. Inicialmente, apresentamos os conjuntos de dados utilizados nos experimentos. Em seguida, detalhamos os procedimentos de pré-treinamento com aprendizado autossupervisionado e de refinamento com aprendizado de poucos exemplos, incluindo as configurações de hiperparâmetros, estratégias de treinamento e recursos computacionais empregados.

5.2.1 Conjuntos de dados

Os experimentos foram conduzidos utilizando dois conjuntos de dados amplamente utilizados na literatura de identificação de medicamentos: NLM PIR e CURE. Para ambos, selecionamos apenas as imagens capturadas em condições não controladas, ou seja, aquelas que não foram obtidas em ambientes laboratoriais com iluminação e orientação padronizadas. Essa escolha é motivada pela necessidade de avaliar a robustez do modelo em cenários do mundo real, onde as imagens de comprimidos podem variar significativamente em termos de iluminação, ângulo e resolução.

Para garantir uma avaliação imparcial, dividimos as classes dos conjuntos de forma disjunta entre treino, validação e teste, evitando qualquer sobreposição entre classes durante os diferentes estágios do processo.

5.2.2 Pré-treinamento com AMDIM

O primeiro estágio do modelo SSL4PILL consiste no pré-treinamento autossupervisionado de um codificador convolucional utilizando o modelo AMDIM. O treinamento foi realizado sobre o conjunto *mini-ImageNet*, amplamente utilizado para aprendizado de representação em tarefas de classificação com poucos exemplos. Esse conjunto contém 100 classes com imagens naturais de diversos objetos, e sua diversidade semântica favorece a aprendizagem de características visuais transferíveis.

Utilizamos uma ResNet-50 como codificador base, com profundidade de 8 camadas convolucionais e largura de canais configurada para 192. A saída do codificador é projetada em um espaço de características de 1536 dimensões, simulando um espaço reprodutor de Hilbert (*Reproducing Kernel Hilbert Space*, RKHS).

Conduzimos o treinamento por 50 épocas com um tamanho de lote de 480 imagens. Como otimizador, utilizamos o *Adam*, com taxa de aprendizado inicial de 2×10^{-4} . O conjunto de transformações aplicadas às imagens inclui recorte aleatório, redimensionamento, oscilação de brilho e contraste, e conversão para tons de cinza, garantindo diversidade nas visualizações geradas para o aprendizado contrastivo.

5.2.3 Ajuste fino com ProtoNet

Após o pré-treinamento, o codificador foi integrado a uma ProtoNet para realizar a classificação de comprimidos em regime de poucos exemplos. O treinamento foi realizado separadamente para os conjuntos de dados NLM PIR e CURE.

Adotamos a configuração clássica de *1-shot 5-way*, com um exemplo de suporte e um exemplo de consulta por classe em cada episódio. Durante o treinamento, foram amostrados 100 episódios por época, com 30 épocas para cada conjunto de dados. Os episódios foram construídos de modo a simular a tarefa real de inferência com categorias não vistas durante o treinamento.

A função de similaridade utilizada foi o quadrado da distância euclidiana, e a escala das similaridades no espaço latente foi ajustada com um parâmetro de temperatura fixado em 128. Para a otimização do modelo com poucos exemplos, utilizamos o algoritmo SGD, com taxa de aprendizado de 1×10^{-4} , reduzida por um fator $\gamma = 0,2$ a cada 10 épocas, favorecendo uma adaptação estável ao longo do tempo.

Durante o treinamento, o codificador pré-treinado foi refinado juntamente com o restante da rede, permitindo que o modelo adaptasse suas representações ao novo domínio visual de forma eficiente.

5.2.4 Ambiente computacional

Todos os experimentos foram implementados com a biblioteca *PyTorch* e executados no ambiente *Google Colab Pro*, utilizando uma única GPU NVIDIA A100.

A execução completa do modelo, incluindo o pré-treinamento com AMDIM e o ajuste com ProtoNet, levou aproximadamente duas horas por conjunto de dados. Essa configuração demonstra a viabilidade prática da abordagem proposta, tanto em termos de custo computacional quanto de reprodutibilidade dos experimentos.

5.3 Avaliação e resultados

Esta seção apresenta a avaliação empírica do modelo SSL4PILL nos conjuntos de dados NLM PIR e CURE. Iniciamos descrevendo os protocolos de avaliação adotados, seguidos pelos resultados quantitativos obtidos com e sem pré-treinamento autossupervisionado. Em seguida, discutimos a influência do aprendizado autossupervisionado no desempenho final do modelo e comparamos os resultados com abordagens anteriores da literatura. Por fim, analisamos limitações observadas e o potencial de generalização da proposta.

5.3.1 Protocolos de avaliação

Para a avaliação do modelo, seguimos a metodologia utilizada em trabalhos relevantes da literatura, como os de [Ling et al. \(2020\)](#), [Li et al. \(2019\)](#) e [Chee et al. \(2024a\)](#). Utilizamos a configuração de *1-shot 5-way*, em que o modelo deve classificar exemplos de consulta com base em apenas um exemplo de suporte por classe.

A avaliação foi realizada em 1000 episódios amostrados aleatoriamente do conjunto de teste, contendo apenas classes não vistas durante o treinamento. Reportamos a acurácia média por episódio e os intervalos de confiança de 95%, obtidos com base no desvio padrão.

As divisões entre conjuntos de treinamento, validação e teste foram construídas com classes disjuntas, garantindo que o modelo fosse avaliado em sua capacidade de generalização para novas categorias de comprimidos, aspecto central na tarefa de aprendizado com poucos exemplos.

5.3.2 Resultados quantitativos

Os resultados estão apresentados na Tabela 9, que compara o desempenho do SSL4PILL com abordagens anteriores e variantes ablatadas da nossa proposta. As maiores acurácias por conjunto de dados estão destacadas em negrito e cor de fundo laranja.

Tabela 9 – Acurácia média (%) na configuração 1-shot 5-way nos conjuntos de dados NLM PIR e CURE. Ling *et al.* (2020) e Chee *et al.* (2024a) não forneceram intervalos de confiança. Destacamos as maiores acurácias em negrito e cor de fundo laranja.

Modelo	Conjuntos de dados	
	NLM	CURE
MS (LING <i>et al.</i> , 2020)	64,20	53,70
CTM, reportado por Ling <i>et al.</i> (2020)	61,20	50,40
MTL, reportado por Ling <i>et al.</i> (2020)	58,70	47,70
Chee <i>et al.</i> (2024a)	–	89,65
SSL4PILL (Sem autossupervisão)	58,02 ± 0,04	46,14 ± 0,03
SSL4PILL (Com autossupervisão)	83,72 ± 0,05	60,78 ± 0,04

Fonte: Autor.

Nos dados da Tabela 9, observa-se que a versão completa do SSL4PILL (com pré-treinamento autossupervisionado) supera todas as abordagens anteriores no conjunto NLM PIR, com uma margem de 19,5 pontos percentuais sobre o melhor modelo anterior (MS). No conjunto CURE, o SSL4PILL também apresenta desempenho superior às abordagens anteriores, embora fique atrás do modelo proposto por Chee *et al.* (2024a), que, por outro lado, não reporta resultados no conjunto NLM PIR, limitando a comparação direta.

5.3.3 Estudo de ablação: impacto do pré-treinamento

Para isolar o impacto do pré-treinamento autossupervisionado, avaliamos também uma variante do modelo treinada diretamente com ProtoNet, sem o estágio de aprendizado autossupervisionado. Essa versão apresenta desempenho inferior em ambos os conjuntos de dados, confirmando que a qualidade das representações iniciais tem papel fundamental no sucesso da adaptação com poucos exemplos.

A diferença entre as variantes com e sem pré-treinamento foi de 25,7 pontos percentuais no conjunto NLM PIR e 14,6 pontos percentuais no conjunto CURE. Esses resultados sugerem que, em domínios com alta similaridade intraclasse e baixo volume de dados rotulados, como a identificação de comprimidos, o aprendizado autossupervisionado fornece representações mais robustas e transferíveis do que o aprendizado direto a partir de poucos exemplos.

5.3.4 Discussão dos resultados

A análise dos resultados obtidos evidencia que a integração do pré-treinamento com AMDIM ao refinamento com ProtoNet oferece ganhos substanciais de desempenho. No conjunto NLM PIR, considerado mais realista por conter imagens capturadas em condições não controladas, o SSL4PILL atingiu acurácia de 83,72%, demonstrando sua robustez frente à variabilidade visual.

A ablação também revela que o aprendizado com poucos exemplos isolado não é suficiente para alcançar desempenho competitivo nesse tipo de tarefa. Na ausência de um bom ponto de partida (pré-treinamento), o modelo sofre para distinguir categorias visuais muito similares, mesmo com técnicas avançadas de aprendizado métrico.

Adicionalmente, observamos que o SSL4PILL apresenta consistência na generalização entre conjuntos com diferentes características (NLM PIR e CURE), o que indica a versatilidade da proposta. Embora não tenha superado o modelo de [Chee et al. \(2024a\)](#) no conjunto de dados CURE, a proposta se destaca por sua capacidade de transferência entre domínios, aspecto importante para aplicações biomédicas reais.

5.3.5 Limitações e generalização

Apesar dos resultados promissores, a abordagem proposta apresenta limitações importantes. O custo computacional do pré-treinamento, embora viável em hardware acessível (como o *Google Colab Pro*), ainda representa uma barreira para aplicações embarcadas.

Além disso, a eficácia do SSL4PILL depende da escolha apropriada de aumentações no estágio autossupervisionado. Estratégias de aumento mal calibradas podem degradar o aprendizado de representações úteis. Estudos futuros podem explorar o impacto dessas escolhas, bem como a utilização de estratégias de aprendizado contínuo para adaptar o modelo a novos comprimidos sem re-treinamento completo.

Por fim, a estrutura do SSL4PILL é agnóstica ao domínio de aplicação, e seu uso pode ser estendido a outras tarefas de identificação de imagens biomédicas de granularidade fina, reforçando o potencial de generalização deste modelo.

6 Conclusões e trabalhos futuros

A identificação precisa e automatizada de comprimidos representa um desafio relevante na intersecção entre saúde e tecnologia, com potencial de reduzir significativamente os erros de medicação e fortalecer a segurança do paciente. Nesta dissertação, abordamos esse problema por meio da concepção, do desenvolvimento e da avaliação de duas metodologias computacionais distintas, que exploram desde abordagens tradicionais de visão computacional até técnicas avançadas de aprendizado profundo. O percurso desta pesquisa demonstrou a viabilidade e a eficácia de diferentes estratégias para a tarefa de identificação de comprimidos.

A primeira contribuição deste trabalho, detalhada no Capítulo 4, concentrou-se na identificação de formas de comprimidos utilizando uma metodologia de baixo custo computacional. Por meio da extração de atributos geométricos simples e da aplicação de classificadores individuais, demonstramos que é possível alcançar desempenho expressivo em imagens capturadas sob condições controladas. Os resultados foram particularmente notáveis para classificadores como o SVM com *kernel* linear, que atingiu acurácia média de 95,71%, sugerindo que o espaço de características gerado pelos atributos de forma é aproximadamente linearmente separável.

Essa abordagem não apenas validou a eficácia de métodos tradicionais para um subproblema específico, mas também se mostrou superior a outros extratores de atributos tradicionais em testes estatísticos. As principais limitações que identificamos foram a dificuldade de generalização para classes com pouca representatividade e o uso de imagens obtidas em condições controladas, desafios que motivaram a exploração de técnicas mais robustas e escaláveis.

A segunda e principal contribuição desta dissertação foi o desenvolvimento da metodologia SSL4PILL, um modelo composto por dois estágios que integra aprendizado autossupervisionado e aprendizado com poucos exemplos para a identificação de comprimidos. Conforme discutimos no Capítulo 5, projetamos essa abordagem para superar o desafio central da escassez de dados rotulados no domínio biomédico. O primeiro estágio consiste em um pré-treinamento autossupervisionado, isto é, uma fase de treinamento inicial em um grande conjunto de imagens não rotuladas, na qual o modelo aprende representações visuais gerais e transferíveis sem depender de anotações humanas. Nessa etapa, utilizou-se o modelo AMDIM para treinar um codificador visual capaz de capturar características robustas e discriminativas. No segundo estágio, a fase de refinamento com o ProtoNet permitiu adaptar eficientemente esse conhecimento à tarefa específica de identificação de comprimidos, mesmo com um número reduzido de exemplos rotulados.

Os resultados quantitativos comprovaram a superioridade da proposta SSL4PILL, que alcançou 83,72% de acurácia no conjunto de dados NLM PIR, superando significativamente as abordagens anteriores. O estudo de ablação reforçou a importância crucial do estágio de pré-treinamento autossupervisionado, evidenciando que a qualidade das representações iniciais é um fator determinante para o sucesso em tarefas de granularidade fina com dados limitados.

Em síntese, este trabalho cumpriu seus objetivos ao propor e validar duas soluções computacionais para a identificação de comprimidos. Partimos de um método simples e interpretável para um problema bem definido (a classificação de formas) até chegar a uma solução sofisticada e eficiente para o desafio mais amplo e realista de reconhecimento visual de comprimidos. As contribuições que apresentamos nesta dissertação possuem implicações diretas para o avanço de sistemas de apoio à decisão clínica, podendo servir de base para o desenvolvimento de ferramentas práticas que auxiliem farmacêuticos, médicos e pacientes na correta administração de medicamentos.

Como direções para trabalhos futuros, destacamos a exploração de técnicas de aprendizado incremental aplicadas ao modelo SSL4PILL, permitindo que o sistema se adapte a novos comprimidos introduzidos no mercado sem a necessidade de um re-treinamento completo. Outra vertente promissora consiste na incorporação de informações multimodais, como a extração das impressões nos comprimidos por meio de OCR, de modo a enriquecer a representação e aprimorar a distinção entre classes visualmente muito semelhantes. Adicionalmente, a otimização do modelo para execução em dispositivos com recursos limitados, como *smartphones*, por meio de técnicas de compressão e destilação de conhecimento, pode ampliar o impacto prático da solução. Por fim, a estrutura independente de domínio (ou agnóstica) da metodologia SSL4PILL permite sua aplicação em outras tarefas de reconhecimento de imagens biomédicas de granularidade fina, reforçando o potencial de generalização desta pesquisa.

Referências

- ANDERSON, T. W. *An Introduction to Multivariate Statistical Analysis*. [S.l.]: Wiley, 1958. Citado na página 43.
- BACHMAN, P.; HJELM, R. D.; BUCHWALTER, W. Learning representations by maximizing mutual information across views. *Advances in Neural Information Processing Systems*, v. 32, 2019. Citado 3 vezes nas páginas 60, 61 e 62.
- BATES, D. W. Preventing medication errors: a summary. *American Journal of Health-System Pharmacy*, Oxford University Press, v. 64, n. 14_Supplement_9, p. S3–S9, 2007. Citado na página 18.
- BATES, D. W.; SINGH, H. Two decades since to err is human: An assessment of progress and emerging priorities in patient safety. *Health Affairs*, v. 37, n. 11, p. 1736–1743, 2018. ISSN 1544-5208. Citado na página 18.
- BEMT, P. M. V. D.; EGBERTS, A. Drug related problems: definitions and classification. *EJHP Practice*, v. 2007, n. 13, p. 62–64, 2007. Citado na página 19.
- BENGIO, Y. Practical recommendations for gradient-based training of deep architectures. In: *Neural Networks: Tricks of the Trade*. [S.l.]: Springer, 2012. p. 437–478. Citado na página 45.
- BISHOP, C. M. *Pattern Recognition and Machine Learning*. 1. ed. Berlin, Heidelberg: Springer-Verlag, 2006. ISBN 9780387310732. Citado 2 vezes nas páginas 81 e 82.
- BRENNAN, T. A. *The Institute of Medicine report on medical errors —- could it do harm?* [S.l.]: Mass Medical Soc, 2000. Citado na página 18.
- BUSTA, M.; NEUMANN, L.; MATAS, J. Deep textspotter: An end-to-end trainable scene text localization and recognition framework. In: *Proceedings of the IEEE international conference on computer vision*. [S.l.: s.n.], 2017. p. 2204–2212. Citado na página 28.
- CABAN, J. J.; ROSEBROCK, A.; YOO, T. S. Automatic identification of prescription drugs using shape distribution models. In: *2012 19th IEEE International Conference on Image Processing*. IEEE, 2012. p. 1005–1008. ISBN 978-1-4673-2533-2. ISSN 1522-4880. Disponível em: <<http://ieeexplore.ieee.org/document/6467032>>. Citado 2 vezes nas páginas 24 e 33.
- CHANG, W.-J. *et al.* Medglasses: A wearable smart-glasses-based drug pill recognition system using deep learning for visually impaired chronic patients. *IEEE Access*, IEEE, v. 8, p. 17013–17024, 2020. Citado 2 vezes nas páginas 32 e 33.
- CHEE, H.-Z. *et al.* Transfer learning for few-shot pill image recognition. In: IEEE. *2024 IEEE 12th Conference on Systems, Process & Control (ICSPC)*. [S.l.], 2024. p. 77–82. Citado 6 vezes nas páginas 11, 29, 33, 66, 67 e 68.
- CHEE, H.-Z. *et al.* Class imbalance pill image classification via ensemble learning. In: *2024 IEEE 12th Conference on Systems, Process & Control (ICSPC)*. [S.l.: s.n.], 2024. p. 41–46. Citado 2 vezes nas páginas 30 e 33.

- CHEN, R.-C. *et al.* An automatic drug image identification system based on multiple image features and dynamic weights. *International Journal of Innovative Computing, Information and Control*, v. 8, p. 2995–3013, 2012. Disponível em: <<http://www.ijicic.org/ijicic-11-01017.pdf>>. Citado 3 vezes nas páginas 23, 24 e 33.
- CHEUNG, K.-C.; BOUVY, M. L.; SMET, P. A. D. Medication errors: the importance of safe dispensing. *British journal of clinical pharmacology*, Wiley Online Library, v. 67, n. 6, p. 676–680, 2009. Citado na página 19.
- CHOLLET, F. Xception: Deep learning with depthwise separable convolutions. In: *2017 IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*. [S.l.: s.n.], 2017. p. 1800–1807. Citado na página 27.
- CHUPAWA, P.; KANJANAWANISHKUL, K. Pill identification with imprints using a neural network. *Maharakham International Journal of Engineering Technology*, p. 30–35, 2015. Citado 2 vezes nas páginas 25 e 33.
- CORTES, C.; VAPNIK, V. Support-vector networks. *Machine Learning*, Springer, v. 20, n. 3, p. 273–297, 1995. Citado 2 vezes nas páginas 43 e 86.
- COSTA, L. d. F.; CESAR JUNIOR, R. M. *Shape Classification and Analysis: Theory and Practice*. 2nd. ed. [S.l.]: Taylor & Francis, 2009. 675 p. Citado 3 vezes nas páginas 40, 41 e 45.
- COVER, T. M.; HART, P. E. Nearest neighbor pattern classification. *IEEE Transactions on Information Theory*, IEEE, v. 13, n. 1, p. 21–27, 1967. Citado na página 43.
- CRAMER, J. A. *et al.* Medication compliance and persistence: terminology and definitions. *Value in health*, Wiley Online Library, v. 11, n. 1, p. 44–47, 2008. Citado na página 20.
- DAILYMED. *Dailymed*. 2025. Disponível em: <<https://dailymed.nlm.nih.gov/dailymed/index.cfm>>. Acesso em: 15 abr. 2025. Citado na página 23.
- DELGADO, N. L. *et al.* Fast and accurate medication identification. *npj Digital Medicine*, v. 2, p. 10, 2 2019. ISSN 2398-6352. Disponível em: <<https://www.nature.com/articles/s41746-019-0086-0>>. Citado 3 vezes nas páginas 23, 27 e 33.
- DENG, J. *et al.* Imagenet: A large-scale hierarchical image database. In: *2009 IEEE Conference on Computer Vision and Pattern Recognition*. [S.l.: s.n.], 2009. p. 248–255. Citado na página 26.
- DIAS, M. L. D. *Novas técnicas baseadas em metaheurísticas para o treinamento de máquinas de vetores suporte*. Mestrado em Ciência da Computação — Instituto Federal de Educação, Ciência e Tecnologia do Ceará (IFCE), Fortaleza, BR, 2017. Citado na página 91.
- DRUGS. *Pill Identifier*. 2025. Disponível em: <https://www.drugs.com/pill_identification.html>. Acesso em: 15 abr. 2025. Citado na página 23.
- DUDA, R. O.; HART, P. E.; STORK, D. G. *Pattern Classification*. 2. ed. New York, NY, US: John Wiley & Sons, 2001. ISBN 9780471056690. Citado 2 vezes nas páginas 43 e 82.
- FISHER, R. A. The use of multiple measurements in taxonomic problems. *Annals of Eugenics*, Wiley Online Library, v. 7, n. 2, p. 179–188, 1936. Citado na página 43.

- GANDHI, T.; SEDER, D.; BATES, D. Methodology matters. identifying drug safety issues: from research to practice. *International Journal for Quality in Health Care*, Oxford University Press, v. 12, n. 1, p. 69–76, 2000. Citado na página 19.
- GONZALEZ, R. C.; WOODS, R. E. *Digital Image Processing (4th Ed)*. [S.l.: s.n.], 2018. 1022 p. ISBN 978-1-292-22304-9. Citado 2 vezes nas páginas 40 e 41.
- GREENE, J. A.; KESSELHEIM, A. S. Why do the same drugs look different? pills, trade dress, and public health. *New England Journal of Medicine*, v. 365, n. 1, p. 83–89, 2011. PMID: 21732842. Citado na página 18.
- HAN, J.; PEI, J.; KAMBER, M. *Data Mining: Concepts and Techniques*. 3rd. ed. Boston, MA, US: Elsevier, 2011. (The Morgan Kaufmann Series in Data Management Systems). ISBN 9789380931913. Citado na página 45.
- HARALICK, R. M.; SHANMUGAM, K.; DINSTEN, I. Textural features for image classification. *IEEE Transactions on Systems, Man, and Cybernetics*, SMC-3, n. 6, p. 610–621, 1973. Citado na página 26.
- HE, K. *et al.* Mask r-cnn. In: IEEE. *Proceedings of the IEEE international conference on computer vision (ICCV)*. [S.l.], 2017. p. 2961–2969. Citado na página 31.
- HE, K. *et al.* Deep residual learning for image recognition. In: *Proceedings of the IEEE conference on computer vision and pattern recognition*. [S.l.: s.n.], 2016. p. 770–778. Citado 2 vezes nas páginas 27 e 29.
- HJELM, R. D. *et al.* Learning deep representations by mutual information estimation and maximization. In: *7th International Conference on Learning Representations, ICLR 2019, New Orleans, LA, USA, May 6-9, 2019*. OpenReview.net, 2019. Disponível em: <<https://openreview.net/forum?id=Bklr3j0cKX>>. Citado na página 60.
- HOWARD, A. G. *et al.* Mobilenets: Efficient convolutional neural networks for mobile vision applications. In: *arXiv preprint arXiv:1704.04861*. [S.l.: s.n.], 2017. Citado na página 27.
- HU, M.-K. Visual pattern recognition by moment invariants. *IRE Transactions on Information Theory*, v. 8, n. 2, p. 179–187, 1962. Citado na página 26.
- HUANG, G.-B.; ZHU, Q.-Y.; SIEW, C. Extreme learning machine: Theory and applications. *Neurocomputing*, Elsevier, v. 70, n. 1-3, p. 489–501, 2006. Citado 2 vezes nas páginas 43 e 86.
- IANDOLA, F. N. *et al.* Squeezenet: Alexnet-level accuracy with 50x fewer parameters and < 0.5 mb model size. *arXiv preprint arXiv:1602.07360*, 2016. Citado na página 27.
- IRVIN, C. B.; ATAS, J. G. Management of evacuee surge from a disaster area: solutions to avoid non-emergent, emergency department visits. *Prehospital and disaster medicine*, Cambridge University Press, v. 22, n. 3, p. 220–223, 2007. Citado na página 20.
- JOCHER, G.; CHAURASIA, A. *YOLOv5 by Ultralytics (Version 7.0)*. 2022. <<https://doi.org/10.5281/zenodo.7347926>>. <<https://zenodo.org/record/7347926>>. Citado na página 31.

- KESSELHEIM, A. S. *et al.* Burden of changes in pill appearance for patients receiving generic cardiovascular medications after myocardial infarction: cohort and nested case-control studies. *Ann Intern Med*, v. 161, n. 2, p. 96–103, 2014. Citado na página 20.
- KESSELHEIM, A. S. *et al.* Variations in pill appearance of antiepileptic drugs and the risk of nonadherence. *JAMA internal medicine*, American Medical Association, v. 173, n. 3, p. 202–208, 2013. Citado na página 20.
- KNERR, S.; PERSONNAZ, L.; DREYFUS, G. Single-layer learning revisited: a stepwise procedure for building and training a neural network. In: SOULIÉ, F. F.; HÉRAULT, J. (Ed.). *Neurocomputing*. Berlin, Heidelberg: Springer Berlin Heidelberg, 1990. p. 41–50. ISBN 978-3-642-76153-9. Citado na página 46.
- KRENZELOK, E. P.; MRVOS, R. The use of an automated interactive voice response system to manage medication identification calls to a poison center. *Clinical Toxicology*, Taylor & Francis, v. 47, n. 5, p. 425–429, 2009. Citado na página 20.
- KRENZELOK, E. P.; MRVOS, R. A regional poison information center ivr medication identification system: Does it accomplish its goal? *Clinical toxicology*, Taylor & Francis, v. 49, n. 9, p. 858–861, 2011. Citado na página 20.
- KRIZHEVSKY, A.; SUTSKEVER, I.; HINTON, G. E. Imagenet classification with deep convolutional neural networks. In: PEREIRA, F. *et al.* (Ed.). *Advances in Neural Information Processing Systems*. Curran Associates, Inc., 2012. v. 25. Disponível em: <https://proceedings.neurips.cc/paper_files/paper/2012/file/c399862d3b9d6b76c8436e924a68c45b-Paper.pdf>. Citado na página 26.
- LEE, Y.-B. *et al.* Pill-id: Matching and retrieval of drug pill images. *Pattern Recognition Letters*, Elsevier, v. 33, p. 904–910, 5 2012. ISSN 0167-8655. Disponível em: <<https://linkinghub.elsevier.com/retrieve/pii/S0167865511002819>>. Citado 2 vezes nas páginas 23 e 33.
- LI, H. *et al.* Finding task-relevant features for few-shot learning by category traversal. In: *Proceedings of the IEEE/CVF conference on computer vision and pattern recognition*. [S.l.: s.n.], 2019. p. 1–10. Citado 2 vezes nas páginas 29 e 66.
- LIN, T.-Y. *et al.* Focal loss for dense object detection. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, v. 42, n. 2, p. 318–327, 2020. Citado na página 30.
- LIN, T.-Y. *et al.* Microsoft coco: Common objects in context. In: FLEET, D. *et al.* (Ed.). *Computer Vision – ECCV 2014*. Cham: Springer International Publishing, 2014. p. 740–755. ISBN 978-3-319-10602-1. Citado na página 31.
- LING, S. *et al.* Few-shot pill recognition. In: *2020 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*. IEEE, 2020. p. 9786–9795. ISBN 978-1-7281-7168-5. ISSN 10636919. Disponível em: <<https://ieeexplore.ieee.org/document/9157392/>>. Citado 10 vezes nas páginas 11, 23, 28, 29, 33, 35, 37, 38, 66 e 67.
- LIU, W. *et al.* Ssd: Single shot multibox detector. In: LEIBE, B. *et al.* (Ed.). *Computer Vision – ECCV 2016*. Cham: Springer International Publishing, 2016. p. 21–37. ISBN 978-3-319-46448-0. Citado na página 30.

- MAKARY, M. A.; DANIEL, M. Medical error — the third leading cause of death in the us. *BMJ*, v. 353, p. i2139, 5 2016. ISSN 1756-1833. Disponível em: <<https://www.bmj.com/lookup/doi/10.1136/bmj.i2139>>. Citado na página 18.
- MCDONALD, C. J.; WEINER, M.; HUI, S. L. Deaths due to medical errors are exaggerated in institute of medicine report. *Journal of the American Medical Association*, v. 284, n. 1, p. 93–95, 2000. ISSN 0098-7484. Citado na página 18.
- MEDICINE, I. of. *To Err Is Human: Building a Safer Health System*. The National Academies Press, 2000. 312 p. ISSN 0273-2300. ISBN 978-0-309-26174-6. Disponível em: <<http://www.nap.edu/catalog/9728>>. Citado na página 18.
- MEDICINE, I. of. *Preventing Medication Errors*. The National Academies Press, 2007. 481 p. ISBN 978-0-309-10147-9. Disponível em: <<http://www.nap.edu/catalog/11623>>. Citado na página 18.
- MEDICINE, N. L. *NLM Pill Image Recognition Challenge 2016*. National Institute of Health, 2019. Disponível em: <<https://pir.nlm.nih.gov/challenge/>>. Acesso em: Jan 15 2019. Citado 3 vezes nas páginas 34, 35 e 36.
- MEDICINE, N. L. *NLM Pill Image Recognition*. National Institute of Health, 2020. Disponível em: <<https://pillbox.nlm.nih.gov/>>. Acesso em: 15 abr. 2025. Citado na página 23.
- NGUYEN, A. D. *et al.* Image-based contextual pill recognition with medical knowledge graph assistance. In: SPRINGER. *Asian Conference on Intelligent Information and Database Systems*. [S.l.], 2022. p. 354–369. Citado 2 vezes nas páginas 30 e 33.
- NGUYEN, A. D. *et al.* High accurate and explainable multi-pill detection framework with graph neural network-assisted multimodal data fusion. *PLOS ONE*, Public Library of Science, v. 18, n. 9, p. 1–27, 09 2023. Disponível em: <<https://doi.org/10.1371/journal.pone.0291865>>. Citado 2 vezes nas páginas 31 e 33.
- NGUYEN, T.-T. *et al.* Multi-stream fusion for class incremental learning in pill image classification. In: *Proceedings of the Asian Conference on Computer Vision*. [S.l.: s.n.], 2022. p. 4565–4580. Citado 2 vezes nas páginas 31 e 33.
- OCHI, S. *et al.* Disaster-driven evacuation and medication loss: a systematic literature review. *PLoS currents*, Public Library of Science, v. 6, 2014. Citado na página 20.
- 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. Citado na página 26.
- OORD, A. van den; LI, Y.; VINYALS, O. Representation learning with contrastive predictive coding. *CoRR*, abs/1807.03748, 2018. Disponível em: <<http://arxiv.org/abs/1807.03748>>. Citado na página 61.
- OU, Y.-Y. *et al.* Automatic drug pills detection based on convolution neural network. In: IEEE. *2018 International Conference on Orange Technologies (ICOT)*. [S.l.], 2018. p. 1–4. Citado 2 vezes nas páginas 27 e 33.

- OZMERMER, T. E. *et al.* Verimedi: Pill identification using proxy-based deep metric learning and exact solution. 4 2021. Disponível em: <<http://arxiv.org/abs/2104.11231>>. Citado 2 vezes nas páginas 31 e 33.
- PLATT, J. Fast training of support vector machines using sequential minimal optimization. In: MIT PRESS. *Advances in Kernel Methods - Support Vector Learning*. Cambridge, MA, US, 1999. p. 21. Citado na página 46.
- RÁDLI, R.; VÖRÖSHÁZI, Z.; CZÚNI, L. Multi-stream pill recognition with attention. In: *2023 IEEE 12th International Conference on Intelligent Data Acquisition and Advanced Computing Systems: Technology and Applications (IDAACS)*. IEEE, 2023. v. 1, p. 942–946. ISBN 979-8-3503-5805-6. ISSN 2770-4254. Disponível em: <<https://ieeexplore.ieee.org/document/10348930/>>. Citado 2 vezes nas páginas 29 e 33.
- RATHE, J. *et al.* Generic switching and non-persistence among medicine users: a combined population-based questionnaire and register study. *PLoS One*, Public Library of Science, v. 10, n. 3, p. e0119688, 2015. Citado na página 21.
- REDMON, J.; FARHADI, A. *YOLOv3: An Incremental Improvement*. 2018. Disponível em: <<https://arxiv.org/abs/1804.02767>>. Citado na página 30.
- REN, S. *et al.* Faster r-cnn: towards real-time object detection with region proposal networks. In: *Proceedings of the 29th International Conference on Neural Information Processing Systems - Volume 1*. Cambridge, MA, USA: MIT Press, 2015. (NIPS'15), p. 91–99. Citado na página 30.
- RUMELHART, D. E.; HINTON, G. E.; WILLIAMS, R. J. Learning representations by back-propagating errors. *Nature*, Nature Publishing Group, v. 323, n. 6088, p. 533–536, 1986. Citado na página 43.
- RXLIST. *NLM Pill Image Recognition*. RxList - The Internet Drug Index for prescription drug information, interactions, and side effects, 2025. Disponível em: <<https://www.rxlist.com>>. Acesso em: 15 abr. 2025. Citado na página 23.
- SMITH, R. An overview of the tesseract ocr engine. In: *Proceedings of the Ninth International Conference on Document Analysis and Recognition - Volume 02*. USA: IEEE Computer Society, 2007. (ICDAR '07), p. 629–633. ISBN 0769528228. Citado na página 24.
- SNELL, J.; SWERSKY, K.; ZEMEL, R. Prototypical networks for few-shot learning. *Advances in neural information processing systems*, v. 30, 2017. Citado na página 60.
- SPILLER, H. A.; GRIFFITH, J. R. Increasing burden of pill identification requests to us poison centers. *Clinical Toxicology*, Taylor & Francis, v. 47, n. 3, p. 253–255, 2009. Citado na página 20.
- STEIN, J. D. *et al.* Impact of the introduction of generic latanoprost on glaucoma medication adherence. *Ophthalmology*, Elsevier, v. 122, n. 4, p. 738–747, 2015. Citado na página 21.
- SUN, Q. *et al.* Meta-transfer learning for few-shot learning. In: *Proceedings of the IEEE/CVF conference on computer vision and pattern recognition*. [S.l.: s.n.], 2019. p. 403–412. Citado na página 29.

- SUNTRONSUK, S.; RATANOTAYANON, S. Pill image binarization for detecting text imprints. In: *2016 13th International Joint Conference on Computer Science and Software Engineering (JCSSE)*. IEEE, 2016. p. 1–6. ISBN 978-1-5090-2033-1. Disponível em: <http://ieeexplore.ieee.org/document/7748859>. Citado na página 24.
- SUNTRONSUK, S.; RATANOTAYANON, S. Automatic text imprint analysis from pill images. In: *2017 9th International Conference on Knowledge and Smart Technology (KST)*. IEEE, 2017. p. 288–293. ISBN 978-1-4673-9077-4. Disponível em: <http://ieeexplore.ieee.org/document/7886081>. Citado 2 vezes nas páginas 24 e 33.
- SZEGEDY, C. *et al.* Rethinking the inception architecture for computer vision. In: *Proceedings of the IEEE conference on computer vision and pattern recognition (CVPR)*. [S.l.: s.n.], 2016. p. 2818–2826. Citado na página 27.
- TAN, L. *et al.* Comparison of retinanet, ssd, and yolo v3 for real-time pill identification. *BMC medical informatics and decision making*, Springer, v. 21, p. 1–11, 2021. Citado na página 30.
- TAN, L. *et al.* Comparison of yolo v3, faster r-cnn, and ssd for real-time pill identification. *Research Square*, 2021. Citado 2 vezes nas páginas 30 e 33.
- TAN, M.; LE, Q. Efficientnet: Rethinking model scaling for convolutional neural networks. In: PMLR. *International conference on machine learning*. [S.l.], 2019. p. 6105–6114. Citado na página 30.
- TEAGUE, M. R. Image analysis via the general theory of moments. *Journal of the Optical Society of America*, v. 70, n. 8, p. 920–930, 1980. Citado na página 26.
- UDE, M. *et al.* Generic switch after ramipril patent expiry is not associated with decreased pharmacy refill compliance: a retrospective study using the dapi database. *Journal of hypertension*, LWW, v. 29, n. 9, p. 1837–1845, 2011. Citado na página 20.
- USHIZIMA, D. *et al.* Investigating pill recognition methods for a new national library of medicine image dataset. In: _____. Springer International Publishing, 2015. v. 9475, p. 410–419. ISBN 978-3-319-27862-9. Disponível em: http://link.springer.com/10.1007/978-3-319-27863-6_38. Citado 2 vezes nas páginas 25 e 33.
- USUYAMA, N. *et al.* epillid dataset: a low-shot fine-grained benchmark for pill identification. In: *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition Workshops*. [S.l.: s.n.], 2020. p. 910–911. Citado 2 vezes nas páginas 28 e 33.
- VIEIRA NETO, M. A. *et al.* CoforDes: An invariant feature extractor for the drug pill identification. In: *2018 IEEE 31st International Symposium on Computer-Based Medical Systems (CBMS)*. IEEE, 2018. v. 2018-June, p. 30–35. ISBN 978-1-5386-6060-7. ISSN 1063-7125. Disponível em: <https://ieeexplore.ieee.org/document/8417208>. Citado 2 vezes nas páginas 25 e 33.
- VINYALS, O. *et al.* Matching networks for one shot learning. *CoRR*, abs/1606.04080, 2016. Disponível em: <http://arxiv.org/abs/1606.04080>. Citado na página 60.

WANG, C.-Y.; BOCHKOVSKIY, A.; LIAO, H.-Y. M. YOLOv7: Trainable bag-of-freebies sets new state-of-the-art for real-time object detectors. In: *2023 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*. [S.l.: s.n.], 2023. p. 7464–7475. Citado na página 30.

WEBMD. *Pill Identification Tool*. 2025. Disponível em: <<https://www.webmd.com/pill-identification>>. Acesso em: 15 abr. 2025. Citado na página 23.

WIJK, B. L. V. *et al.* Generic substitution of antihypertensive drugs: does it affect adherence? *Annals of Pharmacotherapy*, SAGE Publications Sage CA: Los Angeles, CA, v. 40, n. 1, p. 15–20, 2006. Citado na página 20.

WILCOXON, F. Individual comparisons by ranking methods. *Biom. Bull.*, JSTOR, v. 1, n. 6, p. 80, dez. 1945. Citado na página 44.

WONG, Y. F. *et al.* Development of fine-grained pill identification algorithm using deep convolutional network. *Journal of Biomedical Informatics*, v. 74, p. 130–136, 10 2017. ISSN 1532-0464. Disponível em: <<https://linkinghub.elsevier.com/retrieve/pii/S1532046417302022>>. Citado 2 vezes nas páginas 26 e 33.

YU, J. *et al.* Accurate system for automatic pill recognition using imprint information. *IET Image Processing*, v. 9, p. 1039–1047, 12 2015. ISSN 1751-9667. Disponível em: <<https://onlinelibrary.wiley.com/doi/10.1049/iet-ipr.2014.1007>>. Citado 3 vezes nas páginas 23, 24 e 33.

ZENG, X.; CAO, K.; ZHANG, M. Mobiledeeppill: A small-footprint mobile deep learning system for recognizing unconstrained pill images. In: *Proceedings of the 15th Annual International Conference on Mobile Systems, Applications, and Services*. ACM, 2017. p. 56–67. ISBN 9781450349284. Disponível em: <<https://dl.acm.org/doi/10.1145/3081333.3081336>>. Citado 3 vezes nas páginas 23, 26 e 33.

ZHANG, J. *et al.* A forward and backward compatible framework for few-shot class-incremental pill recognition. *IEEE Transactions on Neural Networks and Learning Systems*, IEEE, 2024. Citado 2 vezes nas páginas 29 e 33.

Apêndices

APÊNDICE A – Classificadores para identificação de formas de comprimidos

A.1 Classificadores baseados em distância

Os classificadores baseados em distância são uma classe de algoritmos que utilizam medidas de similaridade ou dissimilaridade entre amostras para realizar a classificação. Esses métodos são intuitivos e frequentemente eficazes, especialmente em problemas onde os dados podem estar bem separados em um espaço métrico.

A.1.1 *k-Nearest Neighbors* (k -NN)

O algoritmo k -NN é um classificador baseado em distâncias amplamente utilizado por sua simplicidade e eficácia. Ele opera sob a premissa de que amostras semelhantes tendem a estar próximas umas das outras em um espaço métrico.

Dada uma nova amostra $\mathbf{x}$ a ser classificada, o k -NN busca as k amostras mais próximas de $\mathbf{x}$ no conjunto de treinamento com base em uma medida de distância, como a distância euclidiana. A classe atribuída a $\mathbf{x}$ é determinada por maioria de votos entre os k vizinhos mais próximos. Formalmente, seja $N_k(\mathbf{x})$ o conjunto dos k vizinhos mais próximos de $\mathbf{x}$. A predição $\hat{y}$ é dada por:

$$\hat{y} = \arg \max_{y_i} \sum_{\mathbf{x}_j \in N_k(\mathbf{x})} \delta(y_j = y_i),$$

onde $\delta(\cdot)$ é a função indicadora, que retorna 1 quando a condição é verdadeira e 0 caso contrário.

O desempenho do k -NN depende fortemente da escolha da métrica de distância e do valor de k . Valores baixos de k tornam o classificador mais sensível a ruídos, enquanto valores altos podem diluir a influência de amostras localmente relevantes. Em problemas de classificação de comprimidos, o k -NN pode ser eficaz quando atributos visuais (como cor, forma e impressão) são representados em um espaço vetorial adequado.

A.1.2 *Linear Discriminant Analysis* (LDA)

O LDA é um classificador que assume que os dados de cada classe seguem uma distribuição normal multivariada com mesma matriz de covariância, mas com médias

distintas. Sob essas hipóteses, ele busca uma projeção linear dos dados que maximize a separabilidade entre as classes, utilizando combinações lineares dos atributos de entrada.

Dada uma amostra $\mathbf{x}$, a função discriminante do LDA para a classe i é dada por:

$$g_i(\mathbf{x}) = \mathbf{x}^\top \Sigma^{-1} \boldsymbol{\mu}_i - \frac{1}{2} \boldsymbol{\mu}_i^\top \Sigma^{-1} \boldsymbol{\mu}_i + \log P(y_i),$$

em que $\boldsymbol{\mu}_i$ é o vetor média da classe i , Σ é a matriz de covariância comum às classes e $P(y_i)$ é a probabilidade a priori da classe i (BISHOP, 2006).

Por considerar a estrutura de covariância dos dados e utilizar uma métrica de distância generalizada (distância de Mahalanobis), o LDA é particularmente eficaz quando os atributos apresentam correlação e a suposição de covariância comum é válida.

A.1.2.1 Interpretação geométrica do LDA

No LDA, como todas as classes compartilham a mesma matriz de covariância Σ , o termo quadrático $\mathbf{x}^\top \Sigma^{-1} \mathbf{x}$ é cancelado ao comparar as funções discriminantes entre duas classes. Consequentemente, a fronteira de decisão entre as classes i e j é linear e pode ser expressa como:

$$(\boldsymbol{\mu}_i - \boldsymbol{\mu}_j)^\top \Sigma^{-1} \mathbf{x} = \frac{1}{2} (\boldsymbol{\mu}_i^\top \Sigma^{-1} \boldsymbol{\mu}_i - \boldsymbol{\mu}_j^\top \Sigma^{-1} \boldsymbol{\mu}_j) - \log \frac{P(y_i)}{P(y_j)}.$$

Essa equação define um hiperplano que separa as regiões de decisão das classes no espaço dos atributos. Além de atuar como classificador, o LDA também pode ser usado como método de redução de dimensionalidade supervisionada. Nesse caso, busca-se uma projeção linear que maximize a razão entre a variância entre classes e a variância dentro das classes, definida como:

$$J(\mathbf{w}) = \frac{\mathbf{w}^\top S_B \mathbf{w}}{\mathbf{w}^\top S_W \mathbf{w}},$$

onde S_B é a matriz de dispersão entre classes e S_W é a matriz de dispersão dentro das classes. A solução ótima corresponde aos autovetores principais de $S_W^{-1} S_B$, resultando em um subespaço de dimensão no máximo $C - 1$, onde C é o número de classes.

A.1.3 Quadratic Discriminant Analysis (QDA)

O classificador QDA estende o classificador LDA ao permitir que cada classe possua sua própria matriz de covariância, relaxando a suposição de homocedasticidade. Dessa forma, o QDA é capaz de modelar fronteiras de decisão não lineares, adequando-se a distribuições com diferentes formas e orientações.

Dada uma amostra $\mathbf{x}$, a função discriminante do QDA para a classe i é dada por:

$$g_i(\mathbf{x}) = -\frac{1}{2} \log |\Sigma_i| - \frac{1}{2} (\mathbf{x} - \boldsymbol{\mu}_i)^\top \Sigma_i^{-1} (\mathbf{x} - \boldsymbol{\mu}_i) + \log P(y_i),$$

onde Σ_i é a matriz de covariância da classe i , $\boldsymbol{\mu}_i$ é seu vetor média e $P(y_i)$ é a probabilidade a priori da classe.

Por utilizar uma métrica de distância adaptada a cada classe, o QDA consegue capturar relações mais complexas entre os atributos. No entanto, requer mais amostras para estimar de forma confiável os parâmetros, o que pode limitar sua aplicabilidade em conjuntos de dados pequenos ou de alta dimensionalidade (BISHOP, 2006).

A.1.3.1 Interpretação geométrica do QDA

No QDA, cada classe possui sua própria matriz de covariância Σ_i , de modo que o termo quadrático $(\mathbf{x}^\top \Sigma_i^{-1} \mathbf{x})$ não é cancelado. Assim, as fronteiras de decisão entre classes são superfícies quadráticas no espaço de características, podendo assumir formas elípticas, parabólicas ou hiperbólicas, dependendo das matrizes de covariância envolvidas.

Geometricamente, o QDA pode ser visto como uma generalização do LDA que permite regiões de decisão curvas e adaptativas. Enquanto o LDA pressupõe que todas as classes compartilhem a mesma forma e orientação (mesma Σ), o QDA admite formas e orientações distintas para cada classe, modelando distribuições mais complexas.

Essa flexibilidade, entretanto, exige a estimação de um número maior de parâmetros ($O(d^2)$ por classe, onde d é a dimensionalidade dos dados), o que torna o QDA mais suscetível a sobreajuste em amostras pequenas.

A.2 Classificadores baseados em probabilidade

A literatura de reconhecimento de padrões (DUDA; HART; STORK, 2001) e aprendizagem de máquina (BISHOP, 2006) apresenta diversas abordagens para o aprendizado em um cenário probabilístico. Suponha que $P(y_i|\mathbf{x})$ denota a probabilidade de uma amostra $\mathbf{x}$ pertencer à classe y_i . A função de custo 0–1, que representa o custo de associar $\mathbf{x}$ à classe incorreta, é minimizada se, e somente se, $\mathbf{x}$ é associada à classe y_k para a qual $P(y_k|\mathbf{x})$ é máxima (DUDA; HART; STORK, 2001). Esse método é conhecido como estimativa *maximum a posteriori* (MAP). Formalmente, a classe associada à amostra $\mathbf{x}$ é dada por:

$$y_{\text{MAP}} = \arg \max_i P(y_i|\mathbf{x}),$$

em que $\arg \max_i$ retorna a classe y_i com maior probabilidade de estar associada a $\mathbf{x}$, ou seja, aquela que maximiza $P(y_i|\mathbf{x})$.

Qualquer função que calcule as probabilidades condicionais $P(y_i|\mathbf{x})$ pode ser interpretada como uma função discriminante, pois permite separar exemplos de classes distintas. Pelo teorema de Bayes, temos:

$$P(y_i|\mathbf{x}) = \frac{P(y_i) P(\mathbf{x}|y_i)}{P(\mathbf{x})}.$$

O denominador $P(\mathbf{x})$ pode ser ignorado, uma vez que é o mesmo para todas as classes e, portanto, não afeta a maximização. Se as probabilidades a priori $P(y_i)$ forem iguais, a decisão pode ser baseada apenas no termo de verossimilhança $P(\mathbf{x}|y_i)$, levando à regra de máxima verossimilhança (MV):

$$h_{MV} = \arg \max_i P(\mathbf{x}|y_i).$$

Embora essa regra seja ótima sob determinadas condições, sua aplicabilidade prática é limitada devido ao grande número de amostras necessárias para estimar de forma confiável as distribuições $P(\mathbf{x}|y_i)$. Para contornar essa limitação, diferentes hipóteses de independência são frequentemente adotadas, resultando em funções discriminantes distintas e, conseqüentemente, em diferentes classificadores. A seguir, é apresentada uma dessas funções, que leva ao classificador *Naïve Bayes*.

A.2.1 *Naïve Bayes* (NB)

O classificador NB baseia-se na hipótese de que os atributos de uma amostra são condicionalmente independentes entre si dado a classe. Sob essa suposição, a verossimilhança $P(\mathbf{x}|y_i)$ pode ser decomposta como:

$$P(\mathbf{x}|y_i) = \prod_{j=1}^d P(x_j|y_i).$$

Assim, a probabilidade posterior de uma amostra pertencer à classe y_i é proporcional a:

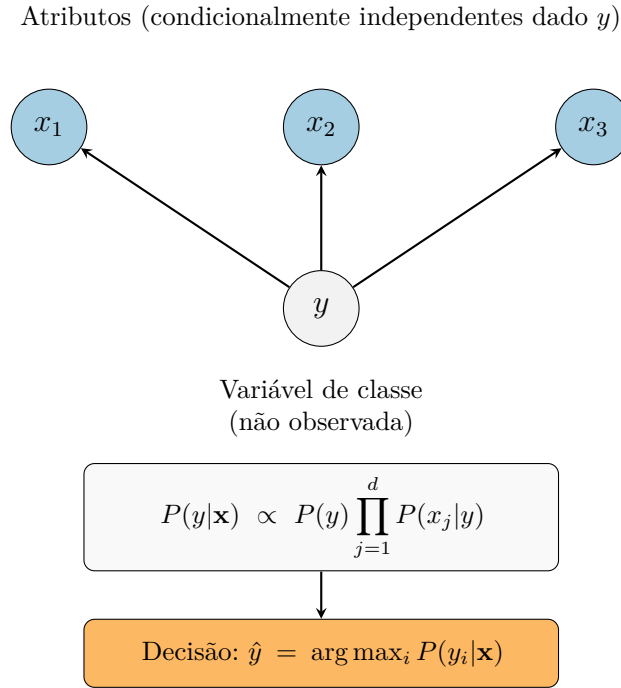
$$P(y_i|\mathbf{x}) \propto P(y_i) \prod_{j=1}^d P(x_j|y_i).$$

O classificador obtido a partir dessa função discriminante, combinado com a regra de decisão MAP, é conhecido como classificador *Naïve Bayes*. O termo *naïve* (ingênuo) refere-se justamente à suposição de independência condicional entre os atributos. A Figura 18 apresenta uma ilustração conceitual do classificador NB.

A fórmula do NB pode ser expressa em uma forma aditiva. Aplicando logaritmo à equação anterior, obtém-se uma forma aditiva mais conveniente para o cálculo numérico:

$$\log P(y_i|\mathbf{x}) = \log P(y_i) + \sum_{j=1}^d \log P(x_j|y_i).$$

Figura 18 – Ilustração conceitual do classificador NB. Cada atributo é modelado por uma distribuição condicional independente, e as probabilidades são combinadas para inferir a classe mais provável.



Fonte: Autor.

A.3 Classificadores baseados em funções

A.3.1 Redes neurais artificiais

Uma rede neural artificial (RNA) é um modelo computacional inspirado no funcionamento do cérebro humano, composto por unidades chamadas “neurônios artificiais”. Uma RNA típica é formada por camadas organizadas em três tipos: camada de entrada, uma ou mais camadas ocultas e camada de saída. Cada neurônio realiza uma combinação linear das entradas recebidas, seguida da aplicação de uma função de ativação não linear, como ReLU, sigmoide ou tangente hiperbólica.

O treinamento de RNAs é realizado por meio de algoritmos de otimização baseados em gradiente, como o gradiente descendente estocástico (*stochastic gradient descent*), utilizando o método de retropropagação do erro (*backpropagation*) para ajustar os pesos sinápticos de modo a minimizar uma função de custo, como o erro quadrático médio ou a entropia cruzada.

Em problemas de classificação de comprimidos, as RNAs podem ser aplicadas diretamente sobre os atributos extraídos das imagens (como forma, cor, textura e impressão) ou em conjunto com técnicas de extração automática de atributos, como nas redes neurais convolucionais (*convolutional neural network*, CNN). Embora redes simples já sejam capazes de aprender padrões complexos, arquiteturas mais profundas tendem a alcançar melhor desempenho em tarefas visuais. No entanto, essas redes demandam maior volume de dados, mais poder computacional e técnicas de regularização (como *dropout*, *batch normalization* e decaimento de pesos) para evitar sobreajuste.

A escolha da arquitetura e dos hiperparâmetros, como número de neurônios por camada e taxa de aprendizagem, influencia diretamente o desempenho do modelo.

A.3.1.1 *Multilayer Perceptron* (MLP)

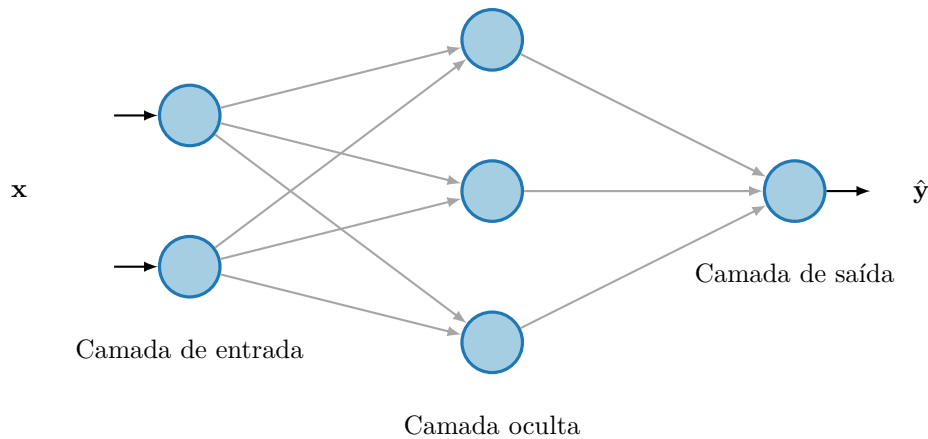
O classificador MLP é uma rede neural *feedforward* composta por uma ou mais camadas ocultas totalmente conectadas. Cada camada realiza uma transformação linear seguida de uma função de ativação não linear. A saída da camada l é dada por:

$$\mathbf{h}^{(l)} = \phi^{(l)}(\mathbf{W}^{(l)}\mathbf{h}^{(l-1)} + \mathbf{b}^{(l)}),$$

em que $\mathbf{W}^{(l)}$ e $\mathbf{b}^{(l)}$ representam, respectivamente, a matriz de pesos e o vetor de vieses da camada l , e $\phi^{(l)}(\cdot)$ é a função de ativação correspondente.

O desempenho do MLP depende diretamente da arquitetura escolhida e dos hiperparâmetros adotados, como taxa de aprendizado, número de épocas e função de ativação. O MLP é adequado para classificação de comprimidos quando há relações não lineares entre os atributos extraídos. A Figura 19 ilustra a arquitetura típica desse classificador.

Figura 19 – Arquitetura típica de um classificador MLP.



Fonte: Autor.

A.3.1.2 *Extreme Learning Machine* (ELM)

O ELM é uma rede neural de camada única com treinamento extremamente rápido. Os pesos entre a camada de entrada e a camada oculta são inicializados aleatoriamente e mantidos fixos durante o treinamento. Apenas os pesos da camada de saída são ajustados, de forma analítica, por meio da solução de um sistema linear.

Seja $\mathbf{H}$ a matriz de ativações da camada oculta e $\mathbf{Y}$ o vetor (ou matriz) de rótulos. A solução ótima para os pesos de saída β é obtida por:

$$\beta = \mathbf{H}^\dagger \mathbf{Y},$$

onde $\mathbf{H}^\dagger$ representa a pseudo-inversa de Moore–Penrose de $\mathbf{H}$ (HUANG; ZHU; SIEW, 2006).

Apesar de sua simplicidade, o ELM pode atingir desempenho competitivo, sendo vantajoso em cenários que exigem baixo tempo de treinamento e boa generalização.

A.3.2 *Support Vector Machine* (SVM)

As SVMs constituem uma das técnicas mais robustas e amplamente utilizadas no campo do aprendizado de máquina supervisionado. Originalmente propostas por Cortes e Vapnik (1995), as SVMs têm como principal objetivo encontrar um hiperplano que separe, da melhor forma possível, os exemplos de duas classes distintas em um espaço de características.

A ideia central do método é maximizar a margem entre os conjuntos de amostras pertencentes a classes opostas, ou seja, a distância entre o hiperplano separador e os exemplos mais próximos de cada classe, denominados vetores de suporte. Essa maximização da margem está diretamente associada a uma melhor capacidade de generalização do classificador, reduzindo o risco de sobreajuste aos dados de treinamento.

Em sua forma mais simples, a SVM busca um hiperplano linear separador em um espaço de entrada. No entanto, por meio do uso de funções de núcleo (*kernels*), o método pode ser estendido para realizar separações não lineares, projetando os dados em espaços de maior dimensionalidade, onde a separabilidade linear pode ser alcançada. Essa característica confere grande flexibilidade ao modelo, permitindo sua aplicação em uma ampla variedade de problemas, incluindo classificação, regressão e detecção de anomalias.

Além disso, a SVM apresenta uma formulação matemática elegante, baseada em um problema de otimização convexa, cuja solução é globalmente ótima. Essa propriedade contrasta com outros métodos de aprendizado, como redes neurais, que podem sofrer com mínimos locais. A introdução de parâmetros de regularização e de variáveis de folga também possibilita lidar com dados ruidosos e casos em que as classes não são perfeitamente

separáveis. Nas subseções seguintes, apresentamos as formulações da SVM para os casos de margens rígidas e margens suaves, bem como a dedução de seus respectivos problemas primal e dual.

A.3.2.1 Margens Rígidas

Considere um conjunto de treinamento $\mathcal{T} = \{(\mathbf{x}_i, d_i)\}_{i=1}^n$, em que $\mathbf{x}_i \in \mathbb{R}^d$ e $d_i \in \{-1, +1\}$. Supondo que os dados sejam linearmente separáveis, a superfície de decisão é definida por um hiperplano:

$$\mathbf{w}^\top \mathbf{x} + b = 0,$$

em que $\mathbf{w}$ é o vetor de pesos e b o viés (ou termo de deslocamento). Os padrões são corretamente classificados se:

$$d_i(\mathbf{w}^\top \mathbf{x}_i + b) \geq 1, \quad \forall i.$$

A margem de separação entre as duas classes é $2/\|\mathbf{w}\|$. O objetivo da SVM é maximizar essa margem, o que equivale a minimizar $\frac{1}{2}\|\mathbf{w}\|^2$. Assim, o problema de otimização primal é formulado como:

$$\min_{\mathbf{w}, b} \frac{1}{2}\|\mathbf{w}\|^2 \quad \text{sujeito a} \quad d_i(\mathbf{w}^\top \mathbf{x}_i + b) \geq 1, \quad i = 1, \dots, n.$$

Esse é um problema de otimização convexa com restrições lineares. Introduzindo multiplicadores de Lagrange $\alpha_i \geq 0$, a Lagrangiana é dada por:

$$\mathcal{L}(\mathbf{w}, b, \boldsymbol{\alpha}) = \frac{1}{2}\|\mathbf{w}\|^2 - \sum_{i=1}^n \alpha_i [d_i(\mathbf{w}^\top \mathbf{x}_i + b) - 1].$$

As condições de otimalidade de Karush–Kuhn–Tucker (KKT) fornecem:

$$\mathbf{w} = \sum_{i=1}^n \alpha_i d_i \mathbf{x}_i, \quad \sum_{i=1}^n \alpha_i d_i = 0.$$

Substituindo essas expressões na Lagrangiana, obtém-se o problema dual:

$$\begin{aligned} \max_{\boldsymbol{\alpha}} \quad & \sum_{i=1}^n \alpha_i - \frac{1}{2} \sum_{i=1}^n \sum_{j=1}^n \alpha_i \alpha_j d_i d_j \mathbf{x}_i^\top \mathbf{x}_j \\ \text{sujeito a} \quad & \sum_{i=1}^n \alpha_i d_i = 0, \quad \alpha_i \geq 0, \quad i = 1, \dots, n. \end{aligned}$$

A solução ótima $\mathbf{w}_\star$ é esparsa, pois apenas alguns vetores (os vetores de suporte) possuem $\alpha_i > 0$. O viés $b_\star$ pode ser calculado a partir de qualquer vetor de suporte que satisfaça $0 < \alpha_i$:

$$b_\star = d_i - \mathbf{w}_\star^\top \mathbf{x}_i.$$

A.3.2.2 Margens Suaves

Quando os dados não são linearmente separáveis, introduzem-se variáveis de folga $\xi_i \geq 0$, que permitem violações da margem. As restrições passam a ser:

$$d_i(\mathbf{w}^\top \mathbf{x}_i + b) \geq 1 - \xi_i, \quad \xi_i \geq 0.$$

O problema primal é então formulado como:

$$\min_{\mathbf{w}, b, \xi} \frac{1}{2} \|\mathbf{w}\|^2 + C \sum_{i=1}^n \xi_i,$$

em que $C > 0$ é o parâmetro de regularização, controlando o equilíbrio entre maximizar a margem e penalizar violações das restrições.

O problema dual correspondente é dado por:

$$\begin{aligned} \max_{\alpha} \quad & \sum_{i=1}^n \alpha_i - \frac{1}{2} \sum_{i=1}^n \sum_{j=1}^n \alpha_i \alpha_j d_i d_j \mathbf{x}_i^\top \mathbf{x}_j \\ \text{sujeito a} \quad & \sum_{i=1}^n \alpha_i d_i = 0, \quad 0 \leq \alpha_i \leq C. \end{aligned}$$

A função objetivo continua dependendo apenas dos produtos internos $\mathbf{x}_i^\top \mathbf{x}_j$, o que permite a extensão natural para o caso não linear por meio de funções de *kernel*.

O vetor ótimo $\mathbf{w}_\star$ é obtido como antes:

$$\mathbf{w}_\star = \sum_{i=1}^n \alpha_i d_i \mathbf{x}_i,$$

e o viés $b_\star$ é calculado como a média sobre os vetores de suporte que satisfazem $0 < \alpha_i < C$:

$$b_\star = \frac{1}{N_s} \sum_{i \in \mathcal{S}} \left(d_i - \sum_{j=1}^n \alpha_j d_j \mathbf{x}_j^\top \mathbf{x}_i \right),$$

onde $\mathcal{S} = \{i : 0 < \alpha_i < C\}$ e $N_s = |\mathcal{S}|$.

APÊNDICE B – Busca em grade e técnicas de validação cruzada

Durante o desenvolvimento de modelos de aprendizado de máquina, é fundamental estimar sua capacidade de generalização, ou seja, seu desempenho diante de dados não vistos durante o treinamento. Para isso, utilizam-se técnicas de validação cruzada, que dividem o conjunto de dados disponível em múltiplas partições, permitindo a avaliação do modelo com base em diferentes subconjuntos de treino e teste.

Esse processo permite estimar o risco estrutural (ou erro esperado) de um algoritmo de forma mais confiável do que avaliá-lo apenas nos dados de treinamento. Além disso, técnicas de validação cruzada são amplamente utilizadas para a escolha dos melhores hiperparâmetros de um modelo, maximizando seu desempenho por meio de abordagens sistemáticas, como a busca em grade.

As principais técnicas de validação cruzada incluem: *holdout*, *k-fold* e *leave-one-out*, cada uma com suas particularidades e indicações.

B.1 Busca em grade

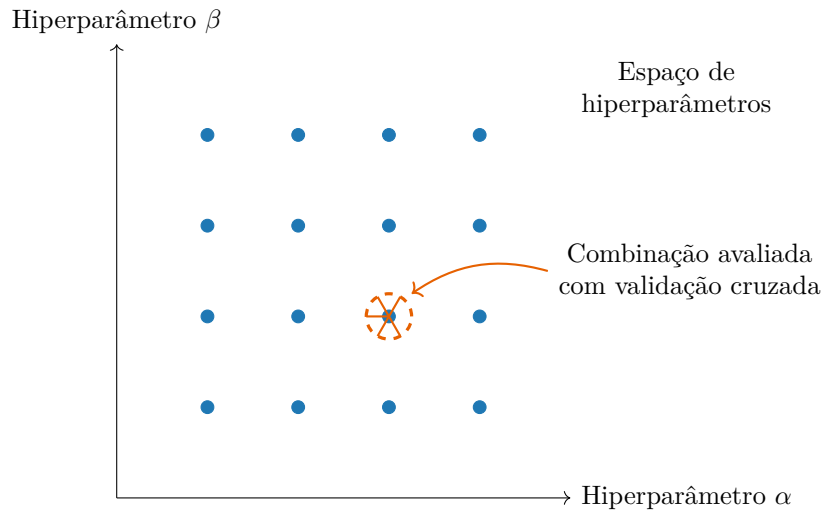
A escolha adequada dos hiperparâmetros de um modelo é uma etapa crítica no processo de modelagem. Para isso, uma técnica amplamente utilizada é a busca em grade (*grid search*), que consiste em definir uma grade (ou malha) de combinações possíveis de hiperparâmetros e avaliá-las sistematicamente por meio de validação cruzada.

Para cada combinação de hiperparâmetros, o modelo é treinado e avaliado dentro de cada fold da validação cruzada, utilizando normalmente a validação cruzada *k-fold* para garantir uma estimativa robusta do desempenho. Em seguida, seleciona-se a combinação que obtiver o melhor desempenho médio segundo uma métrica pré-definida, como acurácia, *f-measure* ou MCC.

A principal vantagem da busca em grade é sua simplicidade e exaustividade, pois garante que todas as combinações possíveis dentro da grade definida sejam testadas. No entanto, essa abordagem pode ser computacionalmente custosa, especialmente quando há muitos hiperparâmetros ou quando os espaços de busca são extensos. Em tais casos, técnicas mais eficientes, como a busca aleatória (*random search*) ou métodos bayesianos, podem ser alternativas viáveis.

A Figura 20 ilustra o processo de busca em grade utilizando validação cruzada.

Figura 20 – Esquema de busca em grade com validação cruzada. Cada ponto da grade representa uma combinação de hiperparâmetros avaliada sistematicamente.



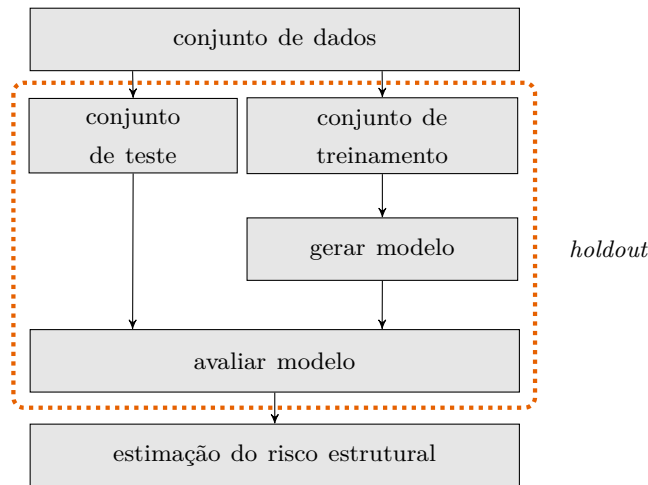
Fonte: Autor.

B.2 Validação cruzada

Técnicas de validação cruzada dividem o conjunto de dados em partições para estimar o erro esperado de um modelo em dados não vistos. Parte dos dados é utilizada para treinar o modelo e a parte restante para avaliação. Esse processo também permite selecionar a combinação de hiperparâmetros que produz o modelo com melhor desempenho para a construção do modelo final.

B.2.1 *Holdout*

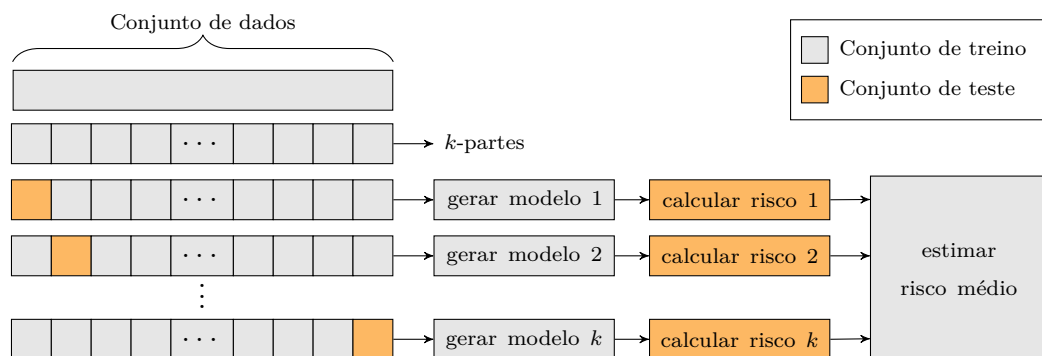
Na técnica *holdout*, o conjunto de dados é dividido em dois subconjuntos mutuamente excludentes: um para treinamento e outro para teste. O desempenho do modelo é avaliado no subconjunto de teste usando métricas como acurácia (Eq. C.1). A Figura 21 ilustra o uso da técnica *holdout* para estimativa do erro esperado.

Figura 21 – Uso da técnica *holdout* para estimação do risco estrutural de um modelo.

Fonte: Adaptação de Dias (2017).

B.2.2 *k-fold*

Na validação cruzada *k-fold*, o conjunto de dados é dividido em k partes mutuamente excluentes e de tamanho aproximadamente igual. Em cada iteração, $k - 1$ partes são usadas para treinar o modelo e a parte restante para validá-lo. O processo é repetido k vezes, alternando a parte de validação. O erro esperado do modelo é então estimado pela média do desempenho nas k iterações. A Figura 22 ilustra a aplicação da técnica *k-fold* para estimação do erro esperado.

Figura 22 – Uso da técnica *k-fold* para estimação do risco estrutural de um modelo.

Fonte: Adaptação de Dias (2017).

APÊNDICE C – Métricas de desempenho

Este apêndice apresenta as principais métricas utilizadas na avaliação de classificadores supervisionados. Essas métricas têm como base a matriz de confusão, uma ferramenta fundamental para a análise quantitativa do desempenho de modelos preditivos.

A matriz de confusão, também chamada de matriz de erros, resume os resultados de um modelo de classificação ao comparar as previsões geradas com os rótulos verdadeiros das amostras. Em sua forma binária, a matriz é composta por quatro categorias: verdadeiros positivos (VP), falsos positivos (FP), verdadeiros negativos (VN) e falsos negativos (FN). As instâncias corretamente classificadas como positivas e negativas são representadas por VP e VN, enquanto FP e FN indicam classificações incorretas.

Embora seja comumente utilizada para problemas de classificação binária, a matriz de confusão pode ser generalizada para cenários com múltiplas classes. Nessa generalização, a matriz assume a forma de uma tabela em que as linhas representam as classes reais e as colunas, as previsões do modelo. As métricas derivadas podem ser adaptadas por meio de abordagens como médias macro, micro ou ponderada.

A seguir, descrevemos as principais métricas derivadas da matriz de confusão, considerando o contexto binário:

- a) a **acurácia (ACC)** mede a proporção de previsões corretas em relação ao total de amostras avaliadas. Embora forneça uma visão geral do desempenho do modelo, pode ser pouco informativa em contextos com desequilíbrio entre as classes;

$$ACC = \frac{VP + VN}{VP + VN + FP + FN} \quad (C.1)$$

- b) a **precisão (PRE)** representa a proporção de instâncias classificadas como positivas que são, de fato, positivas. Essa métrica está relacionada à confiabilidade das previsões positivas;

$$PRE = \frac{VP}{VP + FP} \quad (C.2)$$

- c) a **revocação (REV)**, ou sensibilidade, corresponde à proporção de instâncias positivas corretamente identificadas pelo modelo. É particularmente relevante quando o objetivo é minimizar falsos negativos;

$$REV = \frac{VP}{VP + FN} \quad (C.3)$$

- d) o **coeficiente de correlação de Matthews (MCC)** oferece uma medida global de qualidade da classificação, considerando todas as quatro categorias

da matriz de confusão. Diferentemente da acurácia, o MCC é uma métrica balanceada, adequada para conjuntos de dados desbalanceados.

$$\text{MCC} = \frac{(\text{VP} \cdot \text{VN}) - (\text{FP} \cdot \text{FN})}{\sqrt{(\text{VP} + \text{FP})(\text{VP} + \text{FN})(\text{VN} + \text{FP})(\text{VN} + \text{FN})}} \quad (\text{C.4})$$

Cada métrica fornece uma perspectiva distinta sobre o desempenho de um classificador. A escolha das métricas a serem utilizadas deve considerar o contexto do problema, o balanceamento das classes e os custos associados a diferentes tipos de erro. A análise combinada dessas medidas permite uma avaliação mais abrangente e informativa sobre o comportamento do modelo.

APÊNDICE D – Identificação das imagens utilizadas no Capítulo 3

Este apêndice apresenta a identificação completa dos arquivos de imagem utilizados nas figuras ilustrativas do Capítulo 3. As tabelas a seguir relacionam, para cada figura, os nomes originais dos arquivos correspondentes aos exemplos visuais empregados na descrição dos conjuntos de dados explorados neste trabalho.

Os arquivos referem-se a imagens selecionadas a partir dos conjuntos de dados NLM PIR e CURE, ambos descritos no capítulo supracitado. A identificação explícita dos nomes dos arquivos tem por objetivo garantir a rastreabilidade dos exemplos apresentados, permitindo que outros pesquisadores localizem precisamente as mesmas amostras nas bases originais e, assim, reproduzam ou aprofundem as análises.

Para o conjunto NLM PIR, os nomes de arquivo mantêm o formato original disponibilizado pelo desafio *Pill Image Recognition Challenge*, o qual inclui informações como o código NDC, o tipo de subdivisão do lote (*PART_1_OF_1*) e um identificador de amostra gerado automaticamente. Já no caso do conjunto CURE, as imagens seguem uma estrutura hierárquica de diretórios que especifica a categoria do comprimido, a visão capturada (*top* ou *bottom*) e o tipo de amostra (*Reference* ou *Customer*).

As Figuras 3 e 4 exemplificam a diversidade visual presente no conjunto NLM PIR, incluindo variações de formato, cor e textura entre diferentes medicamentos. As Figuras 5, 6 e 7 ilustram amostras do conjunto CURE, destacando exemplos de referência e consulta, bem como casos considerados difíceis para os métodos de segmentação e classificação. Por fim, a Figura 8 mostra as versões binarizadas de algumas dessas imagens, após a etapa de pré-processamento descrita no Capítulo 3.

Tabela 10 – Nomes dos arquivos utilizados na Figura 3. A posição indica a ordem das imagens, da esquerda para a direita.

Posição	Nome do arquivo
1	00093-7155-56_PART_1_OF_1_CHAL10_SB_A227D15E.jpg
2	1479.jpg
3	927.jpg

Fonte: Autor.

Tabela 11 – Nomes dos arquivos utilizados na Figura 4. A posição indica a ordem das imagens, da esquerda para a direita e de cima para baixo.

Posição	Nome do arquivo
1	00002-4464-30_PART_1_OF_1_CHAL10_SF_B8155C1A.jpg
2	59148-0006-13_PART_1_OF_1_CHAL10_SF_97154BCA.jpg
3	00007-4471-20_PART_1_OF_1_CHAL10_SF_7618BB55.jpg
4	00172-5034-60_PART_1_OF_1_CHAL10_SF_052D0288.jpg
5	00002-3238-30_PART_1_OF_1_CHAL10_SF_361E1B30.jpg
6	00603-5771-21_PART_1_OF_1_CHAL10_SF_38231C78.jpg
7	51672-4032-01_PART_1_OF_1_CHAL10_SF_E52672E3.jpg
8	00093-5125-01_PART_1_OF_1_CHAL10_SF_342D9A0C.jpg
9	68968-3500-01_PART_1_OF_1_CHAL10_SB_581E2C71.jpg
10	00007-4892-20_PART_1_OF_1_CHAL10_SF_BE18DF56.jpg
11	00093-7248-06_PART_1_OF_1_CHAL10_SF_7829BC3D.jpg
12	00093-7425-56_PART_1_OF_1_CHAL10_SF_D727EBFF.jpg

Fonte: Autor.

Tabela 12 – Nomes dos arquivos utilizados na Figura 5. A posição indica a ordem das imagens, da esquerda para a direita.

Posição	Caminho do arquivo	Nome do arquivo
1	0/bottom/Reference/	0_ref_bottom.bmp
2	0/bottom/Customer/	0_bottom_22.jpg
3	0/bottom/Customer/	0_bottom_29.jpg

Fonte: Autor.

Tabela 13 – Nomes dos arquivos utilizados na Figura 6. A posição indica a ordem das imagens, da esquerda para a direita e de cima para baixo.

Posição	Caminho do arquivo	Nome do arquivo
1	53/top/Customer/	53_top_2317.jpg
2	121/top/Customer/	121_top_5289.jpg
3	84/top/Customer/	84_top_3709.jpg
4	89/top/Customer/	89_top_3939.jpg
5	148/top/Customer/	148_top_6520.jpg
6	162/top/Customer/	162_top_7105.jpg
7	8/top/Customer/	8_top_343.jpg
8	71/top/Customer/	71_top_3100.jpg
9	72/top/Customer/	72_top_3148.jpg
10	105/top/Customer/	105_top_4623.jpg
11	116/top/Customer/	116_top_5077.jpg
12	136/top/Customer/	136_top_5951.jpg

Fonte: Autor.

Tabela 14 – Nomes dos arquivos utilizados na Figura 7. A posição indica a ordem das imagens, da esquerda para a direita e de cima para baixo.

Posição	Caminho do arquivo	Nome do arquivo
1	6/top/Customer/	6_top_254.jpg
2	6/top/Customer/	6_top_258.jpg
3	6/top/Customer/	6_top_260.jpg
4	2/top/Customer/	2_top_64.jpg
5	4/top/Customer/	4_top_172.jpg
6	9/top/Customer/	9_top_383.jpg

Fonte: Autor.

Tabela 15 – Nomes dos arquivos utilizados na Figura 8. A posição indica a ordem das imagens, da esquerda para a direita.

Posição	Nome do arquivo
1	00002-3228-30_PART_1_OF_1_CHAL10_SF_391E1C80.jpg
2	00074-6122-90_PART_1_OF_1_CHAL10_SF_051D829C.jpg
3	00006-0112-54_PART_1_OF_1_CHAL10_SF_131609E0.jpg

Fonte: Autor.