



UNIVERSIDADE FEDERAL DO MARANHÃO
UNIVERSIDADE FEDERAL DO PIAUÍ
Doutorado em Ciência da Computação Associação
UFMA/UFPI

Maíla de Lima Claro

**Classificação de Leucemias Utilizando Aumento de Dados,
Transferência de Aprendizado e Combinação de CNNs**

Orientador: Prof. Dr. Rodrigo de Melo Souza Veras
Co-orientador: Prof. Dr. André Macedo Santana

São Luís - MA
Novembro, 2022

Maíla de Lima Claro

Classificação de Leucemias Utilizando Aumento de Dados, Transferência de Aprendizado e Combinação de CNNs

TESE DE DOUTORADO

Tese apresentada como requisito parcial
para obtenção do título de Doutor em Ciência
da Computação, ao Doutorado em Ciência
da Computação, Associação UFMA/UFPI.

Orientador: Prof. Dr. Rodrigo de Melo Souza Veras
Co-orientador: Prof. Dr. André Macedo Santana

São Luís - MA
Novembro, 2022

Ficha gerada por meio do SIGAA/Biblioteca com dados fornecidos pelo(a) autor(a).
Diretoria Integrada de Bibliotecas/UFMA

Claro, Maíla de Lima.

Classificação de Leucemias Utilizando Aumento de Dados,
Transferência de Aprendizado e Combinação de CNNs / Maíla
de Lima Claro. - 2022.

86 p.

Coorientador(a): André Macedo Santana.

Orientador(a): Rodrigo de Melo Souza Veras.

Tese (Doutorado) - Programa de Pós-graduação Doutorado
em Ciência da Computação - Associação UFMA/UFPI,
Universidade Federal do Maranhão, São Luís, 2022.

1. Aprendizado profundo. 2. Classificação de imagens.
3. Comitê. 4. Leucemia. 5. Multilevel. I. Santana,
André Macedo. II. Veras, Rodrigo de Melo Souza. III.
Título.

Maíla de Lima Claro

Classificação de Leucemias Utilizando Aumento de Dados, Transferência de Aprendizado e Combinação de CNNs

A presente Tese de Doutorado foi avaliada e aprovada por banca examinadora composta pelos seguintes membros:

Prof. Dr. Rodrigo de Melo Souza Veras

Orientador

Universidade Federal do Piauí

Prof. Dr. André Macedo Santana

Co-orientador

Universidade Federal do Piauí

Prof. Dra. Fátima Nelsizeuma Sombra de Medeiros

Universidade Federal do Ceará

Prof. Dr. João Manuel Ribeiro da Silva Tavares

Faculdade de Engenharia da Universidade do Porto

Prof. Dr. Kelson Rômulo Aires

Universidade Federal do Piauí

Prof. Dr. Geraldo Braz Junior

Universidade Federal do Maranhão

Prof. Dr. Rodrigo de Melo Souza Veras

Orientador

Prof. Dr. Anselmo Cardoso de Paiva

Coordenador

São Luís - MA, 11 de Novembro de 2022

Aos meus pais, Maria Vilani de Lima Claro e José Claro Neto (in memoriam), por sempre me apoiarem e estarem comigo em todos os momentos, principalmente ao meu pai que me viu iniciando esse sonho mas não estará, presencialmente, vendo esse sonho ser realizado.

Agradecimentos

À Deus, meu Pai bondoso e misericordioso, meu Salvador e Senhor, por seu amor eterno e incondicional.

A toda minha família pelo amor, suporte e incentivo. Em especial aos meus pais, José Claro Neto (*In memoriam*) e Maria Vilani de Lima Claro, pela dedicação e esforço em me proporcionar sempre o melhor e estarem ao meu lado em todos os momentos. Aos meus irmãos Maísa de Lima Claro e Maelson de Lima Claro pelos momentos de descontração e por serem uma força na nossa família.

Agradeço ao meu orientador, Rodrigo Melo de Souza Veras, pela paciência, compreensão, disponibilidade e grande contribuição para realização deste trabalho.

A todos os meus familiares que deram todo o tipo de apoio e que sempre acreditaram em mim e na minha capacidade, em especial aos meus avôs e avós maternos e paternos, tias e tios, e minhas primas.

A todos os amigos do laboratório. Em especial, Luis Henrique da Silva Vogado, Eli-neide Silva dos Santos e Justino Duarte Santos que colaboraram, direta ou indiretamente, para a realização deste trabalho, pelo carinho e aprendizado.

Aos professores do curso de Doutorado em Ciência da Computação pelos ensinamentos e apoio.

À Universidade Federal do Piauí e Universidade Federal do Maranhão pelas oportunidades concedidas e apoio à qualificação profissional e desenvolvimento de pesquisa.

À Coordenação de Aperfeiçoamento de Pessoal de Nível Superior (CAPES) pelo apoio financeiro desta pesquisa no curso de doutorado.

Ao CNPq pelo apoio financeiro e pela oportunidade de realizar um semestre de doutorado sanduíche no exterior na Faculdade de Engenharia da Universidade do Porto. Essa experiência contribuiu bastante para minha formação e meu crescimento como pesquisador, tendo impactado diretamente no desenvolvimento desta tese.

"O homem não teria alcançado o possível se, repetidas vezes, não tivesse tentado o impossível."

(Max Weber)

Resumo

O diagnóstico precoce da leucemia, aumenta significativamente as chances de cura dos pacientes. Mas para isto é preciso ter ciência de qual forma primária o paciente é portador, uma vez que pode se manifestar de duas formas: aguda e crônica, estas ainda se subdividem em mieloide e linfóide. O foco inicial do presente estudo consiste na forma aguda que é um tipo particular de leucemia, onde causa crescimento celular anormal em um curto período de tempo, o que requer um diagnóstico preciso e rápido, para aumentar as chances do tratamento ser bem sucedido. Em sequência, expandimos o foco do trabalho para a classificação em outros tipos de leucemia. Um aliado nestes diagnósticos são os modelos de aprendizagem profunda, que têm sido cada vez mais utilizados em sistemas de diagnóstico médico auxiliados por computadores. Este estudo propõe uma rede neural convolucional (*Convolutional Neural Network* - CNN) chamada *Acute Leukemias Recognition Network* (AlertNet), inspirada em blocos convolucionais da rede Vgg16, mas com camadas densas menores. Para definir os parâmetros da rede AlertNet e suas variações, avaliamos diferentes modelos de CNNs e métodos de ajuste fino usando dezoito conjuntos de dados de imagens, com diferentes características de resolução, contraste, cor e textura, o que totalizou 3.536 imagens. Além desta rede, também utilizamos sete redes pré-treinadas para avaliação e comparação dos resultados. Aplicamos operações de aumento de dados para expandir o conjunto de dados de treinamento e dividimos em cinco cenários de classificações de leucemia: três problemas de classificação binária e dois problemas de classificação multiclasse. Para avaliar a capacidade de generalização das CNNs, aplicamos uma técnica de validação cruzada no conjunto de dados. Os resultados dos experimentos foram promissores, com 96,17% de acerto para a classificação de leucemia aguda e com 94,73% e 94,59% de acertos obtidos pelo multilevel e comitê para o cenário com quatro classes, respectivamente. Métodos como multilevel e comitê foram aplicados neste estudo com o objetivo de melhorar o desempenho das redes. Esses métodos contribuíram para reduzir o erro ou variância nas previsões, o que por sua vez, melhora a precisão do modelo.

Palavras-chave: Classificação de imagens; Aprendizado profundo; Comitê; Leucemia; Multilevel.

Abstract

Early diagnosis of leukemia significantly increases patients' chances of cure. But for this, it is necessary to be aware of which primary form the patient has since it can manifest in two ways: acute and chronic; these are still subdivided into myeloid and lymphoid. The initial focus of the present study is the acute form, a particular type of leukemia that causes abnormal cell growth in a short period, requiring an accurate and rapid diagnosis to increase the chances of successful treatment. In sequence, we expanded the focus of the work to the classification in other types of leukemia. Deep learning models are an ally in these diagnoses, which have been increasingly used in computer-aided medical diagnostic systems developed to detect leukemia. This study proposes a Convolutional Neural Network - CNN) called Acute Leukemias Recognition Network (AlertNet) that was inspired by convolutional blocks of the Vgg16 network but with smaller dense layers. To define the parameters of the AlertNet network and its variations, we evaluated different models of CNNs and fine-tuning methods using eighteen datasets of images with different resolution, contrast, color, and texture characteristics, which totaled 3,536 images. In addition to this network, we also used seven pre-trained networks to evaluate and compare results. We applied data augmentation operations to expand the training dataset and divided it into five leukemia classification scenarios: three binary and two multiclass classification problems. To assess the generalizability of CNNs, we applied a cross-validation technique to the dataset. The results of the experiments were promising, with 96.17% accuracy for the classification of acute leukemia and with 94.73% and 94.59% of correct answers obtained by the multilevel and committee in a scenario with four classes, respectively. Methods such as multilevel and ensemble were applied in this study to improve networks' performance. These methods have contributed to reducing the prediction error and variance, improving the model's accuracy.

Keywords: Image Classification; Deep Learning; Ensemble; Leukemia; Multilevel.

Lista de ilustrações

Figura 1 – Em pacientes com leucemia, o crescimento celular ocorre descontroladamente e há um rápido aumento de leucócitos anormais. Fonte: <i>Cleveland Clinic</i> , 2019.	18
Figura 2 – Exemplos de imagens utilizadas neste trabalho. (a-c) imagens com ALL, (d-f) imagens com AML, (g-i) imagens com leucemia crônica e (j-l) imagens HBS.	20
Figura 3 – Exemplos das imagens utilizadas neste estudo (a-d) imagens HBS, (e-h) imagens ALL, (i-l) imagens AML e (m-p) imagens com outros tipos de leucemia.	33
Figura 4 – Aplicação de aumento de dados. (a) Imagem Original, (b) Rotação, (c) Translação, (d) Espelhamento, (e) <i>Zoom</i> e (f) Cisalhamento.	35
Figura 5 – Operação realizada nas camadas convolucionais.	36
Figura 6 – Operação de <i>Max pooling</i> realizada em um mapa aleatório de características.	37
Figura 7 – Rede Neural padrão (à esquerda) e Rede Neural aplicando <i>dropout</i> (à direita).	38
Figura 8 – Ilustração da arquitetura AlexNet. Fonte: (KRIZHEVSKY; SUTSKEVER; HINTON, 2012).	40
Figura 9 – Ilustração das arquiteturas pré-treinadas Vgg-16 (Esquerda) e Vgg-19 (Direita). Fonte: (SIMONYAN; ZISSERMAN, 2014)	41
Figura 10 – <i>Residual Network</i> com 34 camadas treináveis. Fonte: (HE et al., 2016).	42
Figura 11 – Esquema genérico do funcionamento do Multilevel.	43
Figura 12 – Exemplo da classificação de comitê adotada.	44
Figura 13 – Topologias dos modelos baseados na AlertNet.	49
Figura 14 – Topologia do modelo proposto.	51
Figura 15 – Fluxograma do comitê proposto.	52
Figura 16 – Fluxograma do método de avaliação utilizado para detectar diferentes tipos de leucemia a partir de imagens de lâminas de sangue.	52
Figura 17 – Exemplos de mapas de ativação para lâminas de sangue, (a) imagens com um leucócito, (b) imagens com vários leucócitos. A primeira coluna contém imagens da classe ALL; a segunda coluna representa a classe AML e a terceira a classe HBS.	58
Figura 18 – Comparação dos melhores resultados obtidos pelas configurações de CNNs individuais, multilevel e comitê no cenários (a) Leucemia vs. HBS, (b) ALL vs. HBS, (c) AML vs. HBS, (d) ALL vs. AML vs. HBS and (e) ALL vs. AML vs. HBS vs. outros tipos.	69

Figura 19 – Exemplos de mapas de ativação para imagens de lâminas de sangue -
1º coluna: imagens originais; 2º coluna: Rede Vgg16 e; 3º coluna:
Rede DenseNet121. (a-c) imagens ALL, (d-f) imagens AML, (g-i)
imagens com Leucemia Crônica e (j-l) imagens HBS. 72

Lista de tabelas

Tabela 1 – Resumo dos trabalhos identificados no estado da arte quanto a: descritor(es), classificador, número de bases de dados, imagens utilizadas e acurácia.	30
Tabela 2 – Resumo dos conjuntos de dados de imagens utilizados.	32
Tabela 3 – Matriz de confusão de duas classes (X,Y).	45
Tabela 4 – Nível de exatidão de uma classificação, conforme o valor do índice kappa.	46
Tabela 5 – Resultados obtidos com <i>Shallow Fine Tuning</i> e <i>Deep Fine Tuning</i>	55
Tabela 6 – Resultados obtidos com <i>K-fold cross validation</i> para as variações da AlertNet.	55
Tabela 7 – Resultados alcançados com diferentes comitês formados por variações da rede AlertNet.	56
Tabela 8 – Resultados alcançados com diferentes comitês formados por CNNs pré-treinadas.	57
Tabela 9 – Comparação entre os resultados obtidos pelo método proposto com os obtidos por métodos relacionados.	57
Tabela 10 – Resultados obtidos para leucemia vs. HBS com e sem aumento de dados aplicando <i>K-fold cross validation</i>	59
Tabela 11 – Resultados obtidos pela variações da AlertNet para o cenário leucemia vs. HBS, com e sem aumento de dados aplicando <i>k-fold cross validation</i>	60
Tabela 12 – Resultados obtidos para ALL vs. HBS com e sem aumento de dados aplicando <i>K-fold cross validation</i>	60
Tabela 13 – Resultados obtidos pelas variações da AlertNet para o cenário ALL vs. HBS com e sem aumento de dados aplicando <i>k-fold cross-validation</i>	61
Tabela 14 – Resultados obtidos para AML vs. HBS com e sem aumento de dados aplicando <i>K-fold cross validation</i>	61
Tabela 15 – Resultados obtidos para o cenário AML vs. HBS com e sem aumento de dados aplicando <i>k-fold cross-validation</i>	62
Tabela 16 – Resultados obtidos para ALL vs. AML vs. HBS com e sem aumento de dados aplicando <i>K-fold cross validation</i>	63
Tabela 17 – Resultados obtidos para o cenário ALL vs. AML vs. HBS com e sem aumento de dados aplicando <i>k-fold cross-validation</i>	63
Tabela 18 – Resultados obtidos para ALL vs. AML vs. HBS vs. outros tipos com e sem Aumento de Dados aplicando <i>K-fold cross validation</i>	64
Tabela 19 – Resultados obtidos com <i>K-fold cross validation</i> para os melhores resultados de aumento de dados individuais. (Os melhores valores estão em negrito).	65

Tabela 20 – Resultados obtidos para o cenário ALL vs. AML vs. HBS vs. outros tipos com e sem aumento de dados aplicando <i>k-fold cross-validation</i> .	65
Tabela 21 – Resultados obtidos com <i>K-fold cross validation</i> para os melhores resultados com multilevel. (Os melhores resultados estão em negrito.)	66
Tabela 22 – Resultados obtidos com <i>K-fold cross validation</i> para os melhores resultados com comitê.	67
Tabela 23 – Comparação entre os resultados obtidos pelas configurações multilevel e comitê propostas com os obtidos pelos métodos relacionados. . .	70

Lista de abreviaturas e siglas

ACS	<i>American Cancer Society</i>
AG	<i>Algoritmo Genético</i>
ALL	<i>Acute Lymphoblastic Leukemia</i>
AML	<i>Acute Myeloid Leukemia</i>
ASH	<i>American Society of Hematology</i>
CAD	<i>Computer Aided Diagnosis</i>
CLL	<i>Chronic Lymphocytic Leukemia</i>
CML	<i>Chronic Myeloid Leukemia</i>
CNNs	<i>Convolutional Neural Networks</i>
DA	Aumento de dados
DFT	<i>Deeply Fine Tuning</i>
DSC	<i>Depth-wise Separable Convolution</i>
ECA	<i>Efficient Channel Attention</i>
FC	<i>Fully-Connected</i>
FP	Falso Positivo
FN	Falso Negativo
GAN	<i>Generative Adversarial Network</i>
GD	Gradiente Descendente
HBS	<i>Healthy Blood Slides</i>
IDH	Índice de Desenvolvimento Humano
ILSVRC	<i>ImageNet Large Scale Visual Recognition Challenge</i>
INCA	Instituto Nacional do Câncer
MLP	<i>MultiLayer Perceptron</i>

ReLu	<i>Rectified Linear units</i>
RNAs	Redes Neurais Artificiais
ROI	<i>Region of Interest</i>
SFT	<i>Shallow Fine Tuning</i>
SGD	<i>Stochastic Gradient Descent</i>
SVM	<i>Support Vector Machine</i>
VP	Verdadeiro Positivo
VN	Verdadeiro Negativo
WD	<i>Sem dropout</i>

Sumário

1	INTRODUÇÃO	17
1.1	Contextualização	17
1.2	Objetivo Geral	20
1.3	Objetivos Específicos	21
1.4	Principais Contribuições e Produções Científicas	21
1.5	Organização do Trabalho	22
2	TRABALHOS RELACIONADOS	23
2.1	Leucemia vs. HBS	24
2.2	ALL vs. HBS	25
2.3	AML vs. HBS	27
2.4	CML vs. HBS	28
2.5	ALL vs. AML vs. HBS	28
2.6	ALL vs. AML vs. CLL vs. CML vs. HBS	29
2.7	Considerações Finais	29
3	MATERIAIS E MÉTODOS	31
3.1	Bases de Imagens	31
3.2	Aumento de Dados	33
3.3	<i>Convolutional Neural Networks - CNNs</i>	35
3.3.1	Hiper-parâmetros	38
3.3.2	Arquiteturas e ImageNet	39
3.4	Multilevel	43
3.5	Comitê de Classificadores	44
3.6	Métricas de Avaliação dos Resultados	45
3.7	Considerações Finais	47
4	MÉTODOS PROPOSTOS	48
4.1	Método para detecção de leucemia aguda	48
4.2	Avaliação do impacto de aumento de dados e uma combinação de CNNs na classificação de leucemia	50
4.3	Considerações Finais	53
5	RESULTADOS E DISCUSSÕES	54
5.1	Resultados da detecção de leucemia aguda	54
5.1.1	Discussões	57

5.2	Resultados da avaliação do impacto do aumento de dados e uma combinação de CNNs na classificação de leucemia	58
5.2.1	Classificação Binária	59
5.2.1.1	Cenário 1: Leucemia vs. HBS	59
5.2.1.2	Cenário 2: ALL vs. HBS	59
5.2.1.3	Cenário 3: AML vs. HBS	61
5.2.2	Classificação Multiclasse	62
5.2.2.1	Cenário 4: ALL vs. AML vs. HBS	62
5.2.2.2	Cenário 5: ALL vs. AML vs. HBS vs. outros tipos	63
5.2.3	Multilevel e Comitê	66
5.2.4	Discussões	68
5.3	Considerações Finais	73
6	CONCLUSÕES E TRABALHOS FUTUROS	74
6.1	Conclusões	74
6.2	Trabalhos Futuros	76
	REFERÊNCIAS	77

1 Introdução

Esta tese apresenta novas arquiteturas de Redes Neurais Convolucionais (*Convolutional Neural Networks* - CNNs) denominadas *Acute Leukemias Recognition Network* (AlertNet) e suas variações para detecção dos tipos de leucemia. Inicialmente o foco do estudo é a leucemia aguda e subsequentemente são avaliados e testados outros tipos de classificações de leucemias utilizando imagens de lâminas de sangue. Para o entendimento sobre o tema abordado esse capítulo apresenta o problema, os objetivos da tese, bem como as principais contribuições alcançadas e a produção científica.

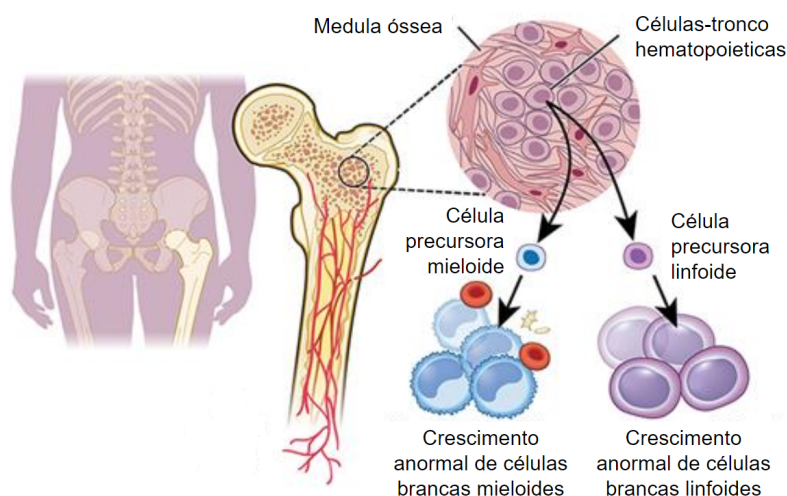
1.1 Contextualização

De acordo com [Soerjomataram e Bray \(2021\)](#), o câncer é atualmente o primeiro ou o segundo contribuinte mais comum para a mortalidade prematura na maioria dos países do mundo. O número global de pacientes com câncer deve aumentar nos próximos 50 anos devido à forte influência das mudanças demográficas, como o envelhecimento e o crescimento da população. Presumindo que as tendências de incidência mais recentes continuem para os principais tipos de câncer, prevê-se uma duplicação da incidência de todos os cânceres combinados até 2070 em relação a 2020. Os maiores aumentos são previstos em ambientes com menos recursos, em países atualmente com baixo Índice de Desenvolvimento Humano (IDH), enquanto os aumentos previstos na carga nacional diminuem com o aumento dos níveis de IDH nacional. Um dos tipos de cânceres mais prevalentes na atualidade é a leucemia, sendo ela uma das doenças que afetam diretamente o funcionamento da medula óssea ([TRAVLOS, 2006](#)).

A medula óssea é o local de fabricação das células sanguíneas e ocupa a cavidade dos ossos. Nela são encontradas as células que dão origem aos glóbulos vermelhos (hemácias ou eritrócitos), às plaquetas e aos glóbulos brancos (leucócitos) ([SOUZA; GORINI, 2006](#)). Estes últimos participam ativamente do sistema imunológico humano e auxiliam na defesa do corpo contra os invasores. As células progenitoras na medula, também conhecidas como células estaminais ou células precursoras, produzem uma média de 100 milhões de leucócitos por dia. Esses auxiliam no combate e eliminação de microrganismos e estruturas químicas estranhas ao organismo por meio de sua captura (fagocitose) ou produção de anticorpos.

A leucemia é uma doença maligna dos glóbulos brancos, geralmente, de origem desconhecida. Tem como principal característica o acúmulo de células doentes na medula óssea, que substituem as células sanguíneas normais. Por conta da leucemia, uma célula sanguínea que ainda não atingiu a maturidade sofre uma mutação genética que a

Figura 1 – Em pacientes com leucemia, o crescimento celular ocorre descontroladamente e há um rápido aumento de leucócitos anormais. Fonte: *Cleveland Clinic*, 2019.



transforma em uma célula cancerosa, como pode ser observado na Figura 1. Essa célula anormal não funciona de forma adequada, multiplica-se mais rápido e tem um tempo de vida menor do que as células normais. Dessa forma, as células sanguíneas saudáveis da medula óssea vão sendo substituídas por células anormais cancerosas (SAWYERS; DENNY; WITTE, 1991).

A *American Cancer Society* (ACS)¹ estima 60.650 novos casos de leucemia para 2022, tendo aproximadamente 24.000 mortes. Ainda de acordo com a ACS, 35.810 casos serão em homens e 24.840 em mulheres, com 14.020 mortes em homens e 9.980 em mulheres.

Os tipos de leucemia podem ser agrupados de acordo com a velocidade e o agravamento da doença. Sob esse aspecto, a doença pode ser do tipo crônica (que geralmente agrava-se lentamente) ou aguda (que costuma piorar de maneira rápida). Elas também podem ser agrupadas baseando-se nos tipos de glóbulos brancos que afetam (linfóides ou mielóides) (MROZEK; HEEREMA; BLOOMFIELD, 2004). Sendo assim, os principais tipos de leucemia são a *Acute Lymphoblastic Leukemia* (ALL), *Acute Myeloid Leukemia* (AML), *Chronic Myeloid Leukemia* (CML) e *Chronic Lymphocytic Leukemia* (CLL). A leucemia aguda atinge, principalmente, crianças e a leucemia crônica tende a atingir adultos e idosos (SOUZA; GORINI, 2006). Existem várias formas de se realizar a detecção precoce da leucemia. A principal delas é a investigação por meio de exames clínicos, laboratoriais e radiológicos em pessoas que já apresentam sinais ou sintomas que indicam a leucemia.

O diagnóstico geralmente passa por um exame de imagens geradas por *Hematology analyzers*, que são equipamentos utilizados para realizar hemogramas, e que

¹ <https://cancerstatisticscenter.cancer.org/#/cancer-site/Leukemia>

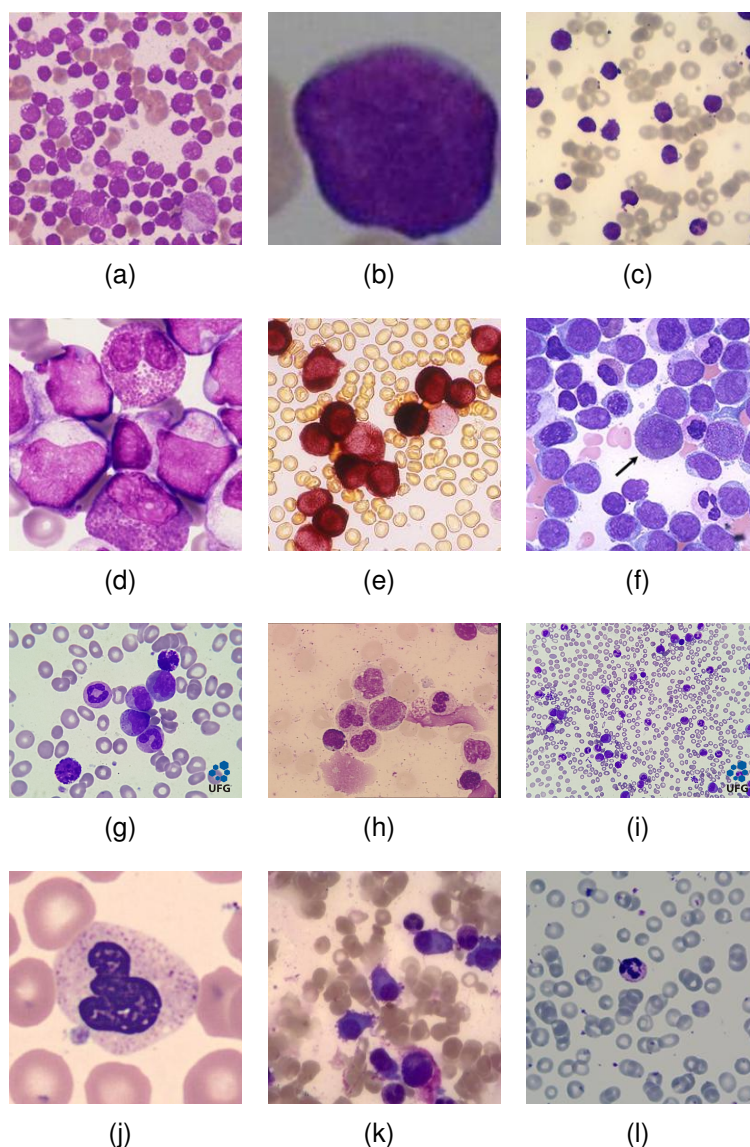
a partir destes exames serão analisadas as informações específicas sobre os tipos e quantidades dos componentes no sangue, como: glóbulos vermelhos (hemácias), glóbulos brancos (leucócitos) e plaquetas (coagulação sanguínea). A classificação das células geralmente depende de suas características morfológicas e requer um médico qualificado. Este procedimento pode ser demorado, tedioso e caro (DWIVEDI, 2018). Além disso, a contagem às vezes pode levar a erros na classificação incorretas de leucócitos (PATEL; WEINBERG, 2020). Assim, vários trabalhos vêm desenvolvendo técnicas automatizadas baseadas em aprendizado de máquina que podem economizar um tempo precioso e reduzir significativamente o esforço e o erro humano (AGAIAN; MADHUKAR; CHRONOPOULOS, 2018).

Cada tipo de leucemia tem um tratamento adequado de acordo com suas especificações, portanto, elas requerem um diagnóstico nos estágios iniciais da doença para fornecer esse tratamento apropriado. Para estágios mais avançados, os principais tratamentos buscam destruir as células leucêmicas para que a medula óssea volte a produzir células normais. A Figura 2 mostra exemplos de imagens de lâmina de sangue usadas em nossos experimentos com ALL, AML, leucemia crônica e lâminas de sangue saudável (*Healthy Blood Slides* - HBS).

Devido à importância clínica em geral, muitos pesquisadores investem no uso de técnicas de processamento de imagens e aprendizado de máquina para desenvolver métodos computacionais que auxiliem especialistas na interpretação de imagens médicas (ANILKUMAR; MANOJ; SAGI, 2021; SASHANK; JAIN; VENKATESWARAN, 2021; AHMED; NAYAK, 2021). Essas técnicas computacionais são incorporadas em sistemas de Diagnóstico Auxiliados por Computador (*Computer Aided Diagnosis* - CAD) e são aplicadas a vários tipos de imagens médicas para fornecer suporte na análise de doenças (HEMELINGS et al., 2021; ZHAO; JIANG; QIU, 2021; FUHAD et al., 2020). Esses sistemas contribuem para reduzir a carga de trabalho do especialista e aumentar a eficiência da análise. Portanto, são úteis no processo de detecção e diagnóstico de leucemia e os seus tipos, por meio da pré-análise das informações fornecidas pelas imagens de lâminas de sangue.

Modelos de aprendizagem profunda têm sido cada vez mais usados em sistemas CAD (PUTTAGUNTA; RAVI, 2021). Particularmente, as Redes Neurais Convolucionais (*Convolutional Neural Networks* - CNNs) podem aprender representações hierárquicas, desde as características mais gerais nas primeiras camadas convolucionais, até as mais semânticas nas últimas camadas. Atualmente, as CNNs são uma das técnicas mais eficazes usadas em diagnósticos baseados em imagens médicas (UPRETI et al., 2021). Pesquisadores têm buscado expandir a generalização das CNNs usando o aumento de dados e a combinação de CNNs em multilevel (SZEGEDY et al., 2015) e comitê (DIETTERICH, 2000). O método multilevel propõe concatenar as características de duas

Figura 2 – Exemplos de imagens utilizadas neste trabalho. (a-c) imagens com ALL, (d-f) imagens com AML, (g-i) imagens com leucemia crônica e (j-l) imagens HBS.



ou mais CNNs e combiná-las, formando assim uma nova rede. Já o método comitê atua na classificação, onde ao invés de ter somente um classificador para rotular, os comitês combinam diferentes classificadores para produzir um resultado mais eficaz. Todos esses conceitos serão abordados posteriormente neste trabalho.

1.2 Objetivo Geral

O objetivo geral desse trabalho consiste em propor sistemas de auxílio ao diagnóstico capazes de identificar imagens de lâminas de sangue com leucemia. Inicialmente, esse sistemas detectam automaticamente os dos dois tipos de leucemia aguda (ALL-AML), e posteriormente, outros tipos de classificação dos tipos de leucemia nas imagens de lâminas de sangue. Como prosseguimento do trabalho, é realizado um

estudo e avaliação de técnicas que são amplamente utilizadas em sistemas CAD, baseados em CNNs para a detecção automática de leucemia e seus tipos em imagens digitais de lâminas de sangue: principalmente, aumento de dados e configurações de comitê e multilevel (KIM; CHO; CHO, 2021).

1.3 Objetivos Específicos

Para atingir o objetivo geral dessa pesquisa, alguns objetivos específicos precisam ser alcançados:

- Construir CNNs com uma menor quantidade de parâmetros para detectar os tipos de leucemia em cinco cenários de classificação;
- Aplicar a técnica aumento de dados no treinamento das redes, confirmando a eficiência das técnicas de aumento de dados para a capacidade de generalizar modelos de aprendizado profundo;
- Selecionar e testar as redes que irão compor as técnicas de multilevel e comitê para detecção de leucemia e seus tipos para melhorar a performance da rede e diminuir a variância dos resultados;
- Validar os métodos propostos por meio de métricas frequentemente utilizadas em trabalhos de imagens médicas;
- Avaliar os métodos propostos por meio da realização de experimentos, usando o maior número de imagens públicas de lâminas de sangue;
- Comparar a rede proposta com as redes pré-treinadas da literatura.

1.4 Principais Contribuições e Produções Científicas

As principais contribuições deste estudo são as seguintes: a utilização de uma diversidade de imagens advindas de dezoito conjuntos de dados públicos para classificação de leucemia; a rede proposta AlertNet para detectar os tipos de leucemia; a detecção e classificação de leucemia em cinco cenários de diferenciação da leucemia e seus tipos; avaliação do desempenho de vários métodos baseados em CNN na classificação de leucemia; avaliação do impacto de múltiplas técnicas de aumento de dados sobre o desempenho de classificação; e, avaliação e comparação das melhorias de configurações baseados em modelo multilevel e comitê.

A produção científica durante o doutorado resultou na publicação e submissão de artigos científicos a periódicos e congressos, além de capítulos de livros. A seguir estão listadas as publicações diretamente relacionadas à pesquisa descrita neste documento:

1. VOGADO, L. H. S.; **CLARO, M. L.**; SANTOS, J. D.; VERAS, R. M. S.. *Deep Learning em Imagens: aplicando CNNs com Keras e TensorFlow*. Escola Regional de Computação Aplicada à Saúde (ERCAS); Encontro Unificado de Computação do Piauí (ENUCOMPI), 2019. 12ªed.. p. 112-135.
2. **CLARO, M. L.**; VOGADO, L. H. S.; SANTOS, J. D.; VERAS, R. M. S.. Utilização de Técnicas de *Data Augmentation* em Imagens: Teoria e Prática. Sociedade Brasileira de Computação, 2020. 8ªed.. p. 47-71.
3. **CLARO, M. L.**; VOGADO, L. H. S.; VERAS, R. M. S.; SANTANA, A. M.; TAVARES, J. M. R. S.; SANTOS, J. D.; MACHADO, V. P.. *Convolution Neural Network Models for Acute Leukemia Diagnosis*. In: *International Conference on Systems, Signals and Image Processing (IWSSIP)*, 2020, Niteroi. *Proceedings of the 2020 International Conference on Systems, Signals and Image Processing (IWSSIP)*, 2020. v. 1. p. 63-68.
4. **CLARO, M. L.**; VOGADO, L. H. S.; VERAS, R. M. S.; SANTANA, A. M.; SANTOS, J. D.; MACHADO, V. P.; TAVARES, J. M. R. S.. *Detecting Acute Leukemia in Blood Slides Images Using a CNNs Ensemble*. *International Journal of Innovative Computing and Applications* (PRINT), 2022. (aceito para publicação)
5. **CLARO, M. L.**; VERAS, R. M. S.; SANTANA, A. M.; VOGADO, L. H. S.; BRAZ JÚNIOR, G.; MEDEIROS, F. N. S.; TAVARES, J. M. R. S.. *Assessing the Impact of Data Augmentation and a Combination of CNNs on Leukemia Classification*. *Information Science* (PRINT), 2022. (publicado)

1.5 Organização do Trabalho

Os demais capítulos deste trabalho foram organizados em:

- O Capítulo 2 apresenta um resumo dos principais trabalhos encontrados na literatura sobre a detecção de leucemia;
- O Capítulo 3 descreve as principais ferramentas e procedimentos computacionais utilizados, tais como: as bases de imagens, aumento de dados, CNNs, técnicas de multilevel e comitê, e por fim, as métricas de avaliação dos resultados;
- O Capítulo 4 mostra o detalhamento do método para detecção e avaliação de técnicas que vem sendo aplicadas pela literatura para detecção dos tipos de leucemia;
- O Capítulo 5 mostra os resultados e apresenta as discussões sobre este estudo;
- Ao final, o Capítulo 6 apresenta as conclusões e sugestões de trabalhos futuros.

2 Trabalhos Relacionados

Realizou-se uma revisão na literatura objetivando identificar quais técnicas estão sendo utilizadas no diagnóstico de leucemia em imagens de lâminas de sangue, para viabilizar a percepção dos descritores e classificadores que estão sendo aplicados. Quanto à metodologia aplicada, identificamos métodos tradicionais (SINGHAL; SINGH, 2016), (MADHUKAR; AGAIAN; CHRONOPOULOS, 2012), (GOUTAM; SAILAJA, 2015), (RAWAT et al., 2017), (LAOSAI; CHAMNONGTHAI, 2018), e métodos que utilizam *Deep Learning* ou aprendizado profundo (THANH et al., 2018), (VOGADO et al., 2018), (LOEY; NAMAN; ZAYED, 2020), (SHAFIQUE; TEHSIN, 2018). Os métodos tradicionais são caracterizados por terem várias etapas, como pré-processamento, segmentação, extração de características e classificação. Os métodos que usam aprendizado profundo, por outro lado, aplicam *Convolutional Neural Networks* (CNNs) em sua metodologia. A ideia por trás dessa categoria é projetar e construir uma CNN que pega a imagem de entrada e retorna a classe de saída sem passar pelas diferentes tarefas desafiadoras envolvidas em métodos tradicionais, como segmentação por região de interesse (*Region of Interest* - ROI) e extração de características.

Essa revisão foi realizada utilizando os bancos de dados Scopus, *Web of Science* e IEEE *Xplore*. As seguintes *strings* de pesquisa foram usadas: “*leukemia acute classification*”, “*white blood cell classification*” e “*blood smear leukemia classification*” nos campos de engenharia e ciência da computação. Como resultado, a primeira *string* retornou 467 artigos, a segunda retornou 427 artigos e a última retornou 101 artigos. Em seguida, analisamos o título e o resumo de todos os artigos retornados a fim de eliminar: documentos repetidos, aqueles com métodos de classificação não automáticos e aqueles cujo objetivo era apenas a segmentação de leucócitos (JIANG; LIAO; DAI, 2003; ABBAS; MOHAMAD, 2014; ARSLAN; OZYUREK; GUNDUZ-DEMIR, 2014; VOGADO et al., 2016; LI et al., 2016).

Devido a crescente evolução dos sistemas CAD na última década, selecionamos os trabalhos publicados após 2012. Se o leitor deseja conhecer abordagens mais antigas, recomendamos a leitura dos artigos de revisões de literatura escritos por Mishra, Majhi e Sa (2016), Anilkumar, Manoj e Sagi (2020) e Khan et al. (2020), onde os autores destes trabalhos abordam e investigam trabalhos anteriores ao ano de 2012. Mishra, Majhi e Sa (2016) realizaram uma análise comparativa dos métodos publicados entre 2002 e 2014. Anilkumar, Manoj e Sagi (2020) apresentaram uma revisão dos trabalhos disponíveis entre 2004 e 2020 na área de processamento de imagens de lâmina de sangue, destacando a detecção automática de leucemia. Khan et al. (2020) investigaram trabalhos que aplicam aprendizado de máquina tradicional e métodos baseados em

aprendizado profundo. Esses autores coletaram 80 artigos publicados em periódicos, livros, conferências e materiais online de 2014 a 2020.

Podemos classificar os artigos encontrados baseado em dois critérios: (1) a metodologia aplicada; e, (2) a quantidade de tipos de leucemia presentes na(s) base(s) de imagens aplicada. Devido à quantidade de trabalhos obtidos fizemos uma divisão em seis abordagens para a detecção de leucemia. Encontramos estudos que realizavam: (1) o diagnóstico entre imagens saudáveis e com leucemia, independente do tipo de leucemia; (2) alguns autores diferenciaram lâminas de sangue com *Acute Lymphocytic Leukemia* (ALL) e lâminas saudáveis; (3) outras propostas diferenciavam as imagens com *Acute Myeloid Leukemia* (AML) e imagens saudáveis; (4) imagens com *Chronic Lymphocytic Leukemia* (CLL) e imagens saudáveis; (5) também foram achados trabalhos que fazem a detecção dos tipos de leucemia aguda (ALL e AML) e imagens saudáveis e, por fim, (6) encontramos estudos que diferenciam os quatro principais tipos de leucemia (ALL, AML, CLL e *Chronic Myeloid Leukemia* (CML)) e imagens saudáveis. A Tabela 1 resume as principais características dos artigos selecionados.

2.1 Leucemia vs. HBS

Para os trabalhos que investigam a classificação binária em imagens de lâminas saudáveis (HBS) ou leucemia, não importando o tipo de leucemia encontrado, foram achados quatro trabalhos. No estudo de Thanh et al. (2018), os autores propuseram um método baseado em CNN para o conjunto de dados ALL-IDB1 (LABATI; PIURI; SCOTTI, 2011) que possui 108 imagens (59 saudáveis e 49 com ALL). Os autores utilizaram técnicas de aumento de dados, tais como imagem refletida, translação, rotação e cisalhamento. O método proposto alcançou 96,6% de acurácia. Em Vogado et al. (2018) os autores utilizaram oito bases de dados: ALL-IDB1, ALL-IDB1 (Crop) ALL-IDB2 (LABATI; PIURI; SCOTTI, 2011), Leukocytes (SARRAFZADEH; DEHNAVI, 2015), CellaVision (ROLLINS-RAVAL; RAVAL; CONTIS, 2012), Altas, Sarrafzadeh et al. (2014) e Sarrafzadeh et al. (2015a). O total de imagens foi de 1.268, onde foram utilizadas as CNNs Vgg-f (CHATFIELD et al., 2014), AlexNet (KRIZHEVSKY; SUTSKEVER; HINTON, 2012) e CaffeNet (JIA et al., 2014) como descritores e o classificador aplicado foi o *Support Vector Machine* (SVM). Os autores alcançaram resultados de 99,76% de acurácia.

O trabalho de Loey, Naman e Zayed (2020) empregou a rede AlexNet (KRIZHEVSKY; SUTSKEVER; HINTON, 2012) para extração e classificação de características. O dataset utilizado nesse trabalho possui 564 imagens (282 saudáveis e 282 com leucemia) e os autores utilizaram as bases de dados *American Society of Hematology* (ASH)¹ e Kangle². Nesse trabalho foram aplicadas técnicas de aumento de dados. Vogado et al. (2021)

¹ <https://imagebank.hematology.org/>

² www.kaggle.com/paultimothymooney/blood-cells

desenvolveram uma rede denominada LeukNet, onde os autores classificaram 3.536 imagens de 18 conjuntos de dados diferentes. Foi obtido uma acurácia de 98,61%. Esse *dataset* utilizado por [Vogado et al. \(2021\)](#) é o mesmo utilizado no presente estudo.

2.2 ALL vs. HBS

Para a segunda abordagem temos a divisão das imagens em ALL e saudáveis. De acordo com o Instituto Nacional do Câncer (INCA)³ a leucemia linfóide aguda afeta células linfóides e agrava-se de maneira rápida. É o tipo mais comum em crianças pequenas, mas também ocorre em adultos. Esse é um dos principais motivos que existem diversos trabalhos que utilizam essa abordagem. No trabalho de [Singhal e Singh \(2016\)](#), os autores extraíram características de textura nas imagens da base de dados ALL-IDB2 ([LABATI; PIURI; SCOTTI, 2011](#)). Nesta base 130 imagens são saudáveis e 130 possuem ALL. Os autores encontraram acurácia de 93,80%. [Shafique e Tehsin \(2018\)](#) aplicaram a rede neural convolucional AlexNet ([KRIZHEVSKY; SUTSKEVER; HINTON, 2012](#)) em um conjunto de dados de 368 imagens das bases de dados ALL-IDB1 e ALL-IDB2 ([LABATI; PIURI; SCOTTI, 2011](#)), totalizando 179 imagens com ALL e 189 imagens saudáveis. Os autores aplicaram aumento de dados no seu estudo e seus resultados registraram 99,50% de acurácia para esta classificação binária. [Ahmed et al. \(2019\)](#) realizaram a classificação em imagens saudáveis e com ALL, onde foi proposta uma nova CNN. Os testes foram aplicados em um total de 179 imagens com ALL e 175 saudáveis e as imagens foram adquiridas dos bancos de dados ASH⁴ e ALL-IDB ([LABATI; PIURI; SCOTTI, 2011](#)). Ainda neste estudo foram aplicadas algumas técnicas de aumento de dados, tais como rotação, translação, espelhamento, cisalhamento e *zoom*, e os autores encontraram uma acurácia de 88,25%.

Ainda na segunda abordagem, temos o trabalho de [Pansombut et al. \(2019\)](#), onde os autores utilizaram o banco de imagens ASH⁵ e o conjunto de dados ALL-IDB1 ([LABATI; PIURI; SCOTTI, 2011](#)) para classificar células em ALL e saudáveis. Os autores utilizaram 121 imagens e também aplicaram técnicas de aumento de dados. ConVNet, a rede proposta pelos autores alcançou 81,74% de acurácia. No trabalho de [Gehlot, Gupta e Gupta \(2020\)](#), os autores desenvolveram uma nova CNN chamada SDCT AuxNet^θ. O *dataset* contém um total de 118 indivíduos ([GUPTA et al., 2019](#)) e os autores utilizaram 15.114 imagens de células segmentadas, obtendo 93,40% de acurácia.

[Das e Meher \(2021\)](#) desenvolveram um modelo híbrido baseado nas redes MobilenetV2 e ResNet18 para detectar lâminas de sangue saudáveis e ALL. Foram utilizados dois conjuntos de dados (ALL-IDB1 e ALL-IDB2) com 368 imagens no total (179

³ <https://www.inca.gov.br/tipos-de-cancer/leucemia>

⁴ <https://imagebank.hematology.org/>

⁵ <https://imagebank.hematology.org/>

com ALL e 189 saudáveis). Dois testes foram realizados com diferentes números de imagens de treinamento e teste. O melhor resultado foi alcançado usando 70% das imagens para treinamento e 30% para teste. Foram obtidas acurácias de 97,92% e 96,00% nos conjuntos de dados ALL-IDB1 e ALL-IDB2, respectivamente.

[Khandekar et al. \(2021\)](#) utilizaram o modelo Yolov4 para detectar as células e classificá-las em ALL e saudáveis. Dois conjuntos de dados foram usados: ALL-IDB1 (108 imagens) e C-NMC-2019. Para o conjunto de dados ALL-IDB1, a validação foi feita em 21 imagens, e 11 foram utilizadas para testar o modelo. As 76 imagens restantes foram usadas para treinamento e, após aumentar as 11 imagens de teste, as células pertencentes a cada classe foram contadas e comparadas com os rótulos do padrão ouro. Já no conjunto de dados C-NMC-2019, o teste foi realizado em 1.000 imagens, sendo 500 imagens pertencentes a cada classe. A precisão média obtida (mAP) foi 96,06% para o conjunto de dados ALL-IDB1 e 98,70% para o conjunto de dados C-NMC-2019.

[Ullah et al. \(2021\)](#) utilizaram o conjunto de dados C-NMC-2019 como entrada para a rede VGG16 e aplicou o módulo *efficient channel attention* (ECA) ([WANG et al., 2021](#)) em cada bloco convolucional desta rede para aumentar ainda mais a relevância das características extraídas. ECA é um módulo de atenção de canal extremamente leve e eficiente, construído com base na imposição da interação entre canais, sem redução de dimensionalidade durante o cálculo da atenção do canal. Os autores compararam os desempenhos do modelo VGG16 com e sem o módulo ECA. Este último mostrou que este mecanismo (ECA) ajuda a melhorar a precisão do modelo utilizado, pois explora a relação entre os canais e obtém uma melhor representação das características. A acurácia obtida foi igual a 91,10%.

A pesquisa de [Karar, Alotaibi e Alotaibi \(2022\)](#) propôs uma estrutura inteligente para classificar leucemias agudas usando imagens de microscopia sanguínea. As amostras de sangue anteriores foram coletadas usando dispositivos digitais sem microscopia e enviadas para um servidor em nuvem. Em seguida, este servidor em nuvem usou um classificador *generative adversarial network* (GAN) para identificar automaticamente leucemia ou saudáveis. O classificador desenvolvido foi avaliado em dois conjuntos de dados públicos: ALL-IDB e ASH. Os autores obtiveram taxas de acerto de 98,67% para a classificação binária (ALL ou saudável) e 95,5% para a classificação multiclasse (ALL, AML e células sanguíneas saudáveis).

[Rodrigues et al. \(2022\)](#) propuseram um modelo híbrido aplicando um algoritmo genético (AG) e uma rede neural convolucional residual (ResNet-50V2) para detectar ALL usando imagens de microscopia disponíveis no conjunto de dados ALL-IDB. O algoritmo genético foi utilizado para encontrar os melhores hiper-parâmetros levando ao maior nível de precisão na classificação. Os resultados mostraram que a otimização com AG melhorou a precisão do classificador, que foi de 98,46%. Foram realizados experimentos

em células cortadas e imagens de microscopia em tamanho real e técnicas de aumento de dados foram aplicadas. O modelo proposto foi testado em dois conjuntos de dados: os conjuntos de dados ALL-IDB1 e ALL-IDB2. Vários testes foram realizados de acordo com dois cenários de classificação: ALL-HBS e ALL-AML-HBS, com a melhor precisão de resultado obtida para o cenário de classificação binária igual a 97% quando a técnica de ajuste foi aplicada à rede Vgg16. A melhor acurácia obtida para a classificação segundo três classes foi de 95%, que foi obtida quando o SVM foi treinado com as características da ResNet50.

Dentre os testes apresentados em [Rastogi, Khanna e Singh \(2022\)](#), foi descrito um experimento de classificação binária utilizando o conjunto de dados ALL-IDB2, onde foram utilizados diferentes classificadores treinados, com o *extra trees classifier* (ETC) obtendo o melhor desempenho de extração de características usando uma rede profunda desenvolvido pelos autores: LeuFeatx. A acurácia relatada foi de 96,15%. Em [Saeed et al. \(2022\)](#) foram propostas as arquiteturas de aprendizado profundo *Multi-Attention EfficientNetV2S* e *EfficientNetB3* utilizando a base de dados C-NMC-2019, com 10.661 imagens de células com as quais 7.272 imagens de câncer obtidas de 47 pacientes com ALL e 3.389 imagens normais obtidas de 26 pessoas saudáveis. Os modelos *Multi-Attention EfficientNetV2S* e *EfficientNetB3* alcançaram 99,73% e 99,25% de acurácia, respectivamente.

2.3 AML vs. HBS

Para a terceira abordagem separamos os trabalhos que classificam as imagens em AML e HBS, e encontramos dois trabalhos. [Madhukar, Agaian e Chronopoulos \(2012\)](#) desenvolveram um sistema de classificação que melhorou o contraste da imagem e extraiu cinco características delas. O classificador SVM foi utilizado para a classificação. Os experimentos foram realizados em 50 imagens do *dataset* ASH⁶ e produziram 93,50% de acurácia. O trabalho de [Goutam e Sailaja \(2015\)](#) utilizou 90 imagens do *dataset* ASH⁷, tendo resultado de 98,00% de acurácia. A partir da quantidade de imagens que cada trabalho adquiriu na base de dados ASH, podemos perceber que esse conjunto de dados é um atlas e que cada trabalho está utilizando um subconjunto diferente de imagens.

([DASARIRAJU; HUO; MCCALLA, 2020](#)) desenvolveram um método para identificação e categorização de AML usando o classificador *Random Forest*. Imagens de leucócitos em pacientes com AML e saudáveis foram obtidas de um conjunto de dados disponível publicamente no *The Cancer Imaging Archive*. Dezesesseis características de cada imagem de glóbulos brancos foram extraídas e as cinco características mais importantes foram usadas na etapa de classificação. O número de imagens de cada tipo

⁶ <https://imagebank.hematology.org/>

⁷ <https://imagebank.hematology.org/>

de leucócito foi de 1.274. O classificador *Random Forest* foi treinado para a detecção e classificação de leucócitos imaturos. O modelo obteve 92,99% de acurácia para detecção e 93,45% de acurácia para classificação de leucócitos imaturos em quatro tipos.

2.4 CML vs. HBS

Na quarta abordagem foi encontrado apenas um estudo para a detecção da leucemia crônica. Um dos motivos é por conta da indisponibilidade de *datasets* públicos para se aplicar no estudo. Assim, tem-se o trabalho de [Khosla e Ramesh \(2018\)](#), que utilizou os conceitos de CNNs para o seu estudo. A leucemia crônica não possui muitas imagens disponíveis publicamente, por isso os autores utilizaram aumento de dados. Havia um total de 67 lâminas e as imagens foram divididas em quatro partes e cada imagem foi rotacionada duas vezes para obter o conjunto de dados aumentado com 536 imagens diferentes. Os autores obtiveram resultados de 97,60% de acurácia.

2.5 ALL vs. AML vs. HBS

Os trabalhos encontrados para essa quinta abordagem tem um maior foco na descoberta de leucemia aguda. [Rawat et al. \(2017\)](#) realizaram a segmentação do núcleo do leucócito em 240 imagens, onde 100 são saudáveis, 60 com ALL e 80 com AML. Em seguida, eles analisaram 331 características de cada núcleo segmentado usando o SVM. O trabalho de [Laosai e Chamnongthai \(2018\)](#) também leva em consideração essas classes. Os testes foram realizados em 500 imagens, 150 do tipo ALL, 150 do tipo AML e 200 do tipo HBS. Essas imagens foram adquiridas no *Ubonratchathani Cancer Hospital and Sunpasit International Hospital, na Thailand*. Segundo os autores, os testes mostraram resultados promissores.

Ainda, nesta abordagem, tem o trabalho de [Tran et al. \(2018\)](#). Neste estudo, os autores propuseram uma CNN para classificar os dois tipos de leucemia aguda e imagens sem leucemia. O sistema desenvolvido foi a rede LeukemiaNet. Os autores utilizaram 108 imagens da base ALL-IDB1 ([LABATI; PIURI; SCOTTI, 2011](#)) e 33 imagens AML da Internet. Foram aplicadas técnicas de aumento de dados e foram obtidos resultados com 97,20% de acurácia. [Ghaderzadeh et al. \(2022\)](#) propuseram um comitê de votação majoritária que combina os quatro modelos (InceptionV3, ResNet-V2, Xception, DesNet121) para classificar as células normais e blásticas do conjunto de dados ISBI-2019. Após o pré-processamento e aumento de dados, o comitê alcançou 98,5% de acurácia.

2.6 ALL vs. AML vs. CLL vs. CML vs. HBS

Na última abordagem temos uma divisão de classificação em cinco classes (ALL - AML - CLL - CML - HBS). No trabalho de [Ahmed et al. \(2019\)](#), citado anteriormente, os autores desenvolveram uma CNN e aplicaram em um total de 903 imagens (179 ALL, 179 AML, 185 CLL, 185 CML e 175 HBS). Estas imagens foram adquiridas dos banco de dados ASH⁸ e ALL-IDB ([LABATI; PIURI; SCOTTI, 2011](#)). Neste trabalho foram aplicadas algumas técnicas de aumento de dados, tais como rotação, translação, espelhamento, cisalhamento e *zoom* e os autores encontraram uma acurácia de 81,74%. Já o estudo de [Bibi et al. \(2020\)](#) aplica a rede neural convolucional DenseNet-121 ([HUANG et al., 2017](#)) em 518 imagens (181 ALL, 55 AML, 38 CLL, 57 CML e 187 HBS). Os autores também aplicaram aumento de dados com objetivo de diminuir o *overfitting* e alcançaram uma acurácia de 99,91%.

2.7 Considerações Finais

Neste capítulo foram apresentados estudos encontrados na literatura que abordam a identificação de leucemia, assim como os seus tipos em imagens de lâminas de sangue. Um resumo dos trabalhos apresentados está demonstrado na Tabela 1. Através desta pesquisa é possível verificar uma maior aplicação de CNNs nos trabalhos mais recentes. Também podemos observar que a maioria dos trabalhos avaliam os seus métodos com poucas diversidades nas imagens, onde estes métodos necessitam de maior robustez e de uma maior quantidade de imagens para obtenção de uma melhor confiabilidade nos seus resultados.

⁸ <https://imagebank.hematology.org/>

Tabela 1 – Resumo dos trabalhos identificados no estado da arte quanto a: descritor(es), classificador, número de bases de dados, imagens utilizadas e acurácia.

Trabalho	Descritor(es)	Classificador	Número de bases de dados	Imagens	Acurácia(%)
Leucemia - HBS					
Thanh et al. (2018)	CNN proposta	CNN	1	108	96,60
Vogado et al. (2018)	CNNs Vgg-f, AlexNet, CaffeNet	SVM	8	1.268	99,76
Loey, Naman e Zayed (2020)	CNN AlexNet	CNN	2	564	100
Vogado et al. (2021)	CNN LeukNet	CNN	18	3.536	98,61
ALL - HBS					
Singhal e Singh (2016)	Textura	SVM	1	260	93,80
Shafique e Tehsin (2018)	CNN AlexNet	CNN	2	368	99,50
Ahmed et al. (2019)	CNN proposta	CNN	2	354	88,25
Pansombut et al. (2019)	CNN ConVNet	CNN	2	363	81,74
Gehlot, Gupta e Gupta (2020)	CNN SDCT AuxNet ^o	CNN	1	15.114	93,40
Das e Meher (2021)	MobileNetV2+ ResNet18	CNN	2	368	97,92 96,00
Khandekar et al. (2021)	Yolov4	Yolov4	2	1.108	mAP:96,06 mAP: 98,70
Ullah et al. (2021)	VGG16-ECA	CNN	1	15.114	91,10
Karar, Alotaibi e Alotaibi (2022)	CNN	GAN	1	368	98,65
Rodrigues et al. (2022)	ResNet-50V2-GA	CNN	1	260	98,46
Abhishek et al. (2022)	Vgg16	CNN	2	608	97,00
Rastogi, Khanna e Singh (2022)	LeuFeatx	ETC	1	260	96,15
Saeed et al. (2022)	Multi-Attention EfficientNetV2S e EfficientNetB3	CNN	1	10.661	99,73 99,25
AML - HBS					
Madhukar, Agaian e Chronopoulos (2012)	Textura	SVM	1	50	93,50
Goutam e Sailaja (2015)	Textura	SVM	1	90	98,00
Dasariraju, Huo e McCalla (2020)	Forma e Cor	Random Forest	1	1.274	92,99
CML - HBS					
Khosla e Ramesh (2018)	CNN proposta	CNN	1	67	97,60
ALL - AML - HBS					
Rawat et al. (2017)	Geometria, cor e textura	GA-SVM	1	240	99,50
Laosai e Chamnongthai (2018)	Forma, distribuição de cores, textura e número de nucleoli	SVM	2	500	99,85
Tran et al. (2018)	CNN LeukemiaNet	CNN	2	141	97,20
Karar, Alotaibi e Alotaibi (2022)	CNN proposta	GAN	2	445	95,50
Abhishek et al. (2022)	ResNet50	SVM	2	608	98,00
Ghaderzadeh et al. (2022)	Comitê InceptionV3, ResNet-V2, Xception, DesNet121	CNN	1	12.528	98,50
ALL - AML - CLL - CML - HBS					
Ahmed et al. (2019)	CNN proposta	CNN	2	903	81,74
Bibi et al. (2020)	CNN DenseNet121	CNN	2	518	99,91

3 Materiais e Métodos

Neste capítulo são apresentados os conceitos necessários para a compreensão das técnicas usadas na elaboração do método proposto. As seções a seguir abordam conteúdos sobre as bases de imagens públicas aplicadas, bem como as técnicas de aumento de dados utilizadas nos treinamentos das redes. Também são apresentadas as arquiteturas de redes que são aplicadas no estudo e que, também serviram como base para a implementação das redes AlertNet e suas variações na classificação dos tipos de leucemia e imagens saudáveis. Além disso, os métodos multilevel e comitê são apresentados, cujo objetivo é aumentar a precisão e diminuir a variância das redes. Por fim, são abordadas as métricas de avaliação comumente utilizadas nos trabalhos da literatura.

3.1 Bases de Imagens

Um dos desafios no processamento de imagens de leucemia é a habilidade de diagnosticar a doença em bases de dados com características distintas. No entanto, vários autores testaram seus sistemas com apenas uma base de dados, ou então nas suas próprias bases. Nesse trabalho utilizamos dezoito bases de imagens com características distintas entre si e advindas de vários lugares do mundo.

O desenvolvimento de uma metodologia robusta para auxiliar no diagnóstico depende dos dados utilizados na sua validação. O principal desafio encontrado no estado da arte está relacionado à aquisição dos conjuntos de dados, uma vez que a maioria deles são privados. Porém, obtivemos dezoito conjuntos de dados públicos com 3.536 imagens para a avaliação do modelo proposto. Na Tabela 2, os conjuntos de dados de imagens utilizados são apresentados com informações sobre o número de imagens por classe, o tamanho das imagens, se possui foco apenas em uma célula e de qual país a base é procedente. As bases de dados ASH¹, leukemia-images², UFG³ e MIDB⁴ foram obtidas dos URLs indicados.

Das bases de imagens listadas na Tabela 2, observamos que, do total de imagens, 1.434 são da classe HBS (40,55%), 881 da classe ALL (24,92%), 978 da classe AML (27,66%) e 243 da classe “outros tipos” (6,87%). Concluimos que o conjunto de imagens está desbalanceado, apresentando um desafio adicional à proposta.

¹ <https://imagebank.hematology.org/> (acessado em 16 de agosto de 2021)

² <http://www.leukemia-images.com/> (acessado em 16 de agosto de 2021)

³ <https://hematologia.farmacia.ufg.br> (acessado em 16 de agosto de 2021)

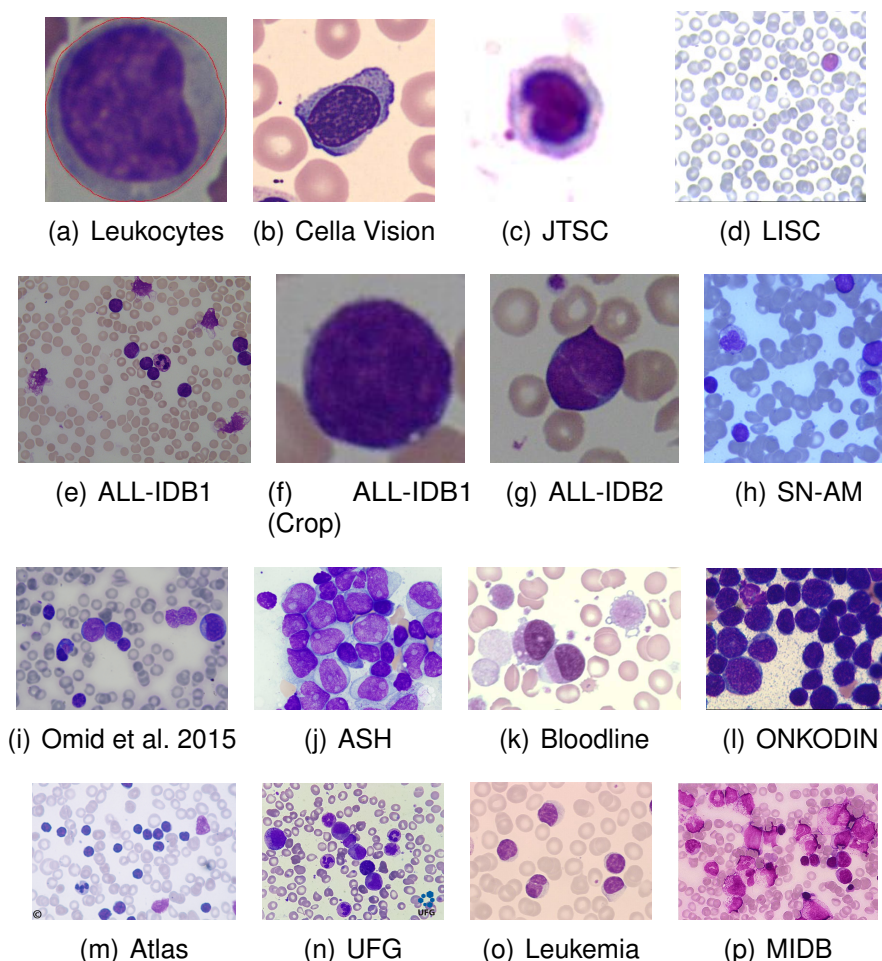
⁴ http://www.midb.jp/blood_db/db.php?lang=en (acessado em 16 de agosto de 2021)

Tabela 2 – Resumo dos conjuntos de dados de imagens utilizados.

Base de Dados	HBS	ALL	AML	outros tipos	Total	Tamanho	Foco na Célula	País	Referência
ALL-IDB 1	59	49	-	-	108	2592 x 1944	Não	Itália	(LABATI; PIURI; SCOTTI, 2011)
ALL-IDB 1 (Crop)	-	510	-	-	510	2592 x 1944	Sim	Itália	(LABATI; PIURI; SCOTTI, 2011)
ALL-IDB 2	130	130	-	-	260	257 x 257	Sim	Itália	(LABATI; PIURI; SCOTTI, 2011)
Leukocytes	149	-	-	-	149	609 x 584	Sim	Irã	(SARRAFZADEH; DEHNAVI, 2015)
CellaVision	109	-	-	-	109	399 x 399	Sim	Suécia	(ROLLINS-RAVAL; RAVAL; CONTIS, 2012)
Atlas	-	25	40	23	88	460x307	Não	-	-
Omid et al. 2014	154	-	-	-	154	3872x2592	Não	Irã	(SARRAFZADEH et al., 2014)
Omid et al. 2015	-	-	27	-	27	3246x2448	Não	Irã	(SARRAFZADEH et al., 2015b)
ASH-OK	-	-	96	-	96	184 x 138	Não	-	link
Bloodline	-	-	217	-	217	1280 x 720	Sim/Não	Brasil	(VALE et al., 2014)
ONKODIN	-	-	78	-	78	768 x 576	Não	Alemanha	(BÖHM, 2008)
CellaVision 2	100	-	-	-	100	300x300	Sim	Suécia	(ZHENG et al., 2018)
JTSC	300	-	-	-	300	2048 x 1536	Sim	China	(ZHENG et al., 2018)
UFG	57	10	27	27	121	640 x 480	Sim/Não	Brasil	link
SN-AM	-	30	-	-	30	224 x 224	Não	Índia	(GUPTA et al., 2019)
leukemia-images	-	40	78	22	140	755x570	Não	-	link
MIDB Dataset	-	87	415	171	673	768x576	Não	Mundo	link
LISC Dataset	376	-	-	-	376	720x576	Não	Irã	(REZATOFIHI; SOLTANIAN-ZADEH, 2011)
Total de imagens	1.434	881	978	243	3.536	-			

Alguns dos dezoito conjuntos de dados coletados têm imagens com apenas um leucócito por imagem e outros têm vários leucócitos por imagem. Apenas os conjuntos de dados UFG e Bloodline possuem os dois tipos de imagens. Assim, a combinação dos conjuntos de dados coletados resultou em um conjunto de dados de imagens com cores, texturas, contrastes e resoluções distintos. Essa diversidade criou um desafio para os modelos em estudo. A Figura 3 mostra exemplos das imagens utilizadas.

Figura 3 – Exemplos das imagens utilizadas neste estudo (a-d) imagens HBS, (e-h) imagens ALL, (i-l) imagens AML e (m-p) imagens com outros tipos de leucemia.



3.2 Aumento de Dados

Overfitting acontece quando um modelo aprende os detalhes e ruídos nos dados de treinamento. Isso significa que o ruído ou flutuações aleatórias nos dados de treinamento são captados e aprendidos como conceitos pelo modelo. O problema é que esses conceitos não se aplicam a novos dados e impactam negativamente a capacidade de generalização dos modelos (SHORTEN; KHOSHGOFTAR, 2019).

Existem algumas técnicas abordadas para a redução do *overfitting*. A maioria

dessas estratégias para aumento da capacidade da generalização se concentram na própria arquitetura do modelo. Soluções como *dropout*, *batch normalization*, *transfer learning* e pré-treinamento foram desenvolvidas para tentar estender o aprendizado profundo para aplicações em conjuntos de dados menores (KUKAČKA; GOLKOV; CREMERS, 2017). Em contraste com essas técnicas, o aumento de dados tenta solucionar o *overfitting* na origem do problema, o conjunto de dados de treinamento.

Dados limitados são um grande obstáculo na aplicação de modelos de aprendizagem profunda. Frequentemente, classes desbalanceadas podem ser um obstáculo adicional, embora possa haver dados suficientes para algumas classes, igualmente importantes, mas classes com poucos dados sofrerão com baixa precisão específica da classe (SHORTEN; KHOSHGOFTAAR, 2019).

O aumento de dados é essencial em métodos baseados em aprendizado profundo, independentemente do problema (LI et al., 2021), particularmente na medicina, para obter segmentações, classificações e análises de imagens. Existem alguns artigos sobre métodos de aumento de dados para tipos específicos de imagens na literatura, como imagens de pele (GOCERI, 2020).

A literatura apresenta vários métodos de aumento de dados, incluindo métodos baseados na semântica de imagens (WANG et al., 2021). No entanto, esses métodos precisam da definição de vários parâmetros de configuração. Assim, decidimos aplicar as formas mais comuns de aumento de dados com base em transformações geométricas. Essas transformações são caracterizadas por sua facilidade de implementação. Isso inclui as transformações de rotação, translação, espelhamento, *zoom* e cisalhamento, Figura 4.

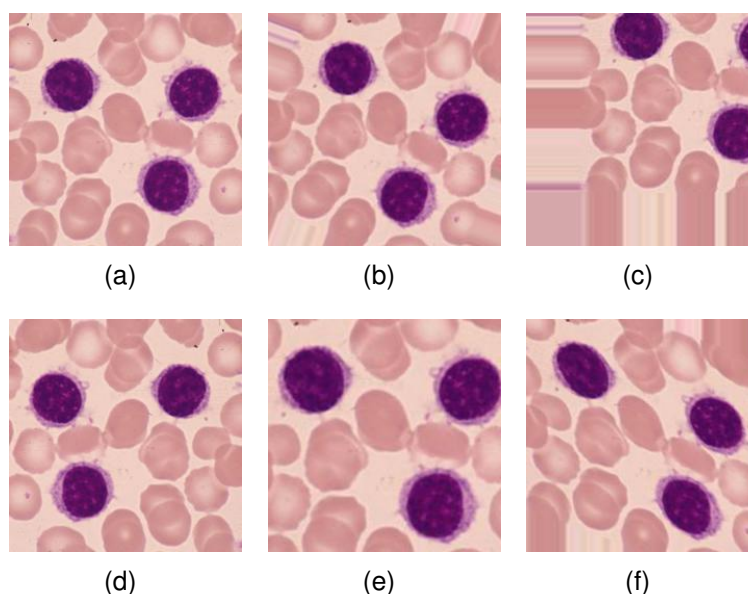
A operação de rotação foi utilizada na imagem em uma direção aleatória (esquerda ou direita). Durante esse processo, os valores de pixel de uma imagem são movidos para a esquerda, direita, para cima e para baixo de acordo com o valor de grau especificado entre 0°-180°, conforme mostrado na Figura 4(b). Um exemplo desta técnica pode ser visto na Figura 4(b).

A translação foi obtida deslocando aleatoriamente os pixels da imagem para cima ou para baixo, para à esquerda ou direita. Por meio dessa transformação, os modelos de aprendizado de máquina não memorizam imagens que estavam constantemente centradas no conjunto de dados. Durante o processo de translação pode haver alterações de altura e largura na imagem, criando espaços sem informação. Uma forma de resolver é preencher esta lacuna com pixels vizinhos. A Figura 4(c) apresenta um exemplo desta técnica.

Na operação de espelhamento os pixels da imagem foram movidos horizontalmente e verticalmente de uma metade da imagem para a outra metade. Foi determinado que os valores dos pixels se moveriam em uma direção aleatória e um exemplo é apresentado na

Figura 4(d). Já a técnica *zoom* aproxima ou afasta a aparência dos objetos na imagem. Essa técnica foi realizada adicionando novos valores de pixel à imagem original. Esses valores dos pixels originais foram examinados e o valor mais próximo foi determinado. Como resultado vemos a Figura 4(e). Também foi utilizado o cisalhamento para deslocar os pixels da imagem no sentido anti-horário de acordo com o ângulo especificado em graus. Neste estudo, esse valor foi definido em 20° e um exemplo é apresentado na Figura 4(f).

Figura 4 – Aplicação de aumento de dados. (a) Imagem Original, (b) Rotação, (c) Translação, (d) Espelhamento, (e) *Zoom* e (f) Cisalhamento.



3.3 Convolutional Neural Networks - CNNs

O processo de extração de características consiste na representação de um conjunto redundante de dados, por meio de características únicas que o façam se diferenciar de outros conjuntos. Existem inúmeras características que podem ser extraídas, dentre elas, as principais em um sistema de auxílio médico, são: cor, forma e textura. Dentre os descritores vistos no estado da arte, a utilização do descritor CNN neste processo de extração de características vem ganhando destaque.

As CNNs fazem parte do conjunto de técnicas de *Deep Learning*. Essas redes convolucionais são uma classe de Redes Neurais Artificiais (RNAs) que modelam abstrações em alto nível por meio de imagens e camadas convolucionais dispostas de forma sequencial ou não. O conceito de CNN foi apresentado por Yann LeCun ([LECUN et al., 1998](#)) e Fukushima ([FUKUSHIMA, 1988](#)) na década de 90. No entanto, apenas no século XXI essa tecnologia foi desenvolvida com eficácia.

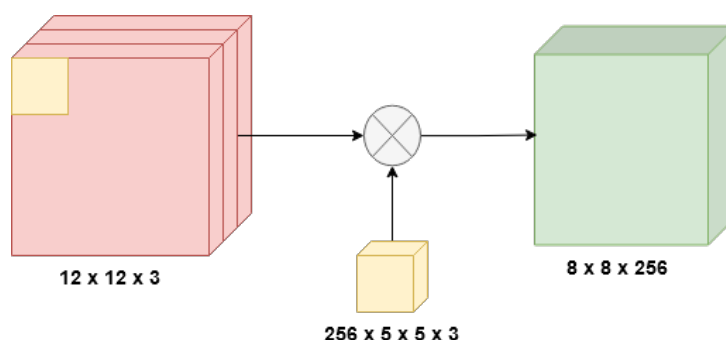
Atualmente as CNNs são empregadas nos mais diversos problemas que envolvem

a classificação, segmentação e detecção em imagens. Uma das principais vantagens na utilização de CNNs é a alta capacidade de aprender as mais diversas características que não são perceptíveis para outras técnicas tradicionais que utilizam descritores. No entanto, essa vantagem demanda um alto custo computacional e a necessidade de grandes bases de dados para o treinamento.

Nos últimos anos, novas arquiteturas vêm surgindo a fim de agregar melhorias em relação às redes anteriores. Tais arquiteturas compartilham entre si os conceitos fundamentais discutidos a seguir.

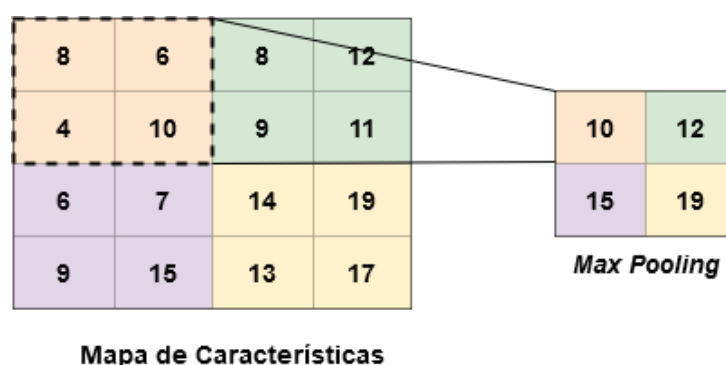
Camadas Convolucionais: Como as demais redes neurais artificiais, as CNNs apresentam estruturas no formato de camadas que auxiliam na extração de características, redução e classificação. As camadas convolucionais são compostas por uma quantidade c de filtros com tamanho $d \times d$ que irão extrair mapas de características ao serem convoluídos com as imagens de entrada ou com saídas de outras camadas. A saída da convolução é denominada mapa de características ou *features map*. Esses mapas geralmente representam características gerais extraídas da imagem, como cor, borda, textura e forma. Um exemplo da operação de convolução é apresentada na Figura 5.

Figura 5 – Operação realizada nas camadas convolucionais.



Camadas de *Pooling*: Após as camadas convolucionais, geralmente são empregadas as de *pooling*. Esse tipo de camada reduz a dimensão espacial dos mapas gerados por camadas anteriores, consequentemente, reduzindo o custo computacional. Dentre os diversos tipos de *pooling* existe o *maxpooling*. Essa operação utiliza uma janela deslizante de tamanho $n \times n$ no mapa de características e para cada passo realizado, o valor máximo daquela janela é retirado. Na Figura 6 essa operação é ilustrada com um mapa aleatório.

Camadas Totalmente Conectadas: A saída das camadas convolucionais e de *pooling* representa as características extraídas da imagem de entrada, que são utilizadas nas camadas totalmente conectadas para a classificação da mesma em uma classe pré-determinada. As camadas densas ou totalmente conectadas (*Fully-Connected layers* - FCs) foram apresentadas inicialmente nas RNAs. São constituídas de neurônios que representam pesos e guardam o aprendizado da rede neural. Nas CNNs elas apresentam

Figura 6 – Operação de *Max pooling* realizada em um mapa aleatório de características.

a mesma função e geralmente aparecem ao final da arquitetura, após inúmeras camadas convolucionais. A operação que transforma os mapas de características em um vetor de neurônios é realizada pela camada *flatten*. A camada de classificação presente nas CNNs é do tipo densa, no entanto, apresenta uma função de ativação do tipo *softmax*, enquanto as FCs contam com ativação por meio da ReLu.

Em resumo, as CNNs são formadas por uma sequência de camadas e cada uma delas possui uma função específica na propagação do sinal de entrada. As camadas convolucionais são responsáveis por extrair atributos dos volumes de entradas. As camadas de *pooling* são responsáveis por reduzir a dimensionalidade do volume resultante após as camadas convolucionais e ajudam a tornar a representação invariante a pequenas translações a entrada. As camadas totalmente conectadas são responsáveis pela propagação do sinal por meio da multiplicação ponto a ponto e o uso de uma função de ativação. A saída da CNN é a probabilidade da imagem de entrada pertencer a uma das classes para a qual a rede foi treinada.

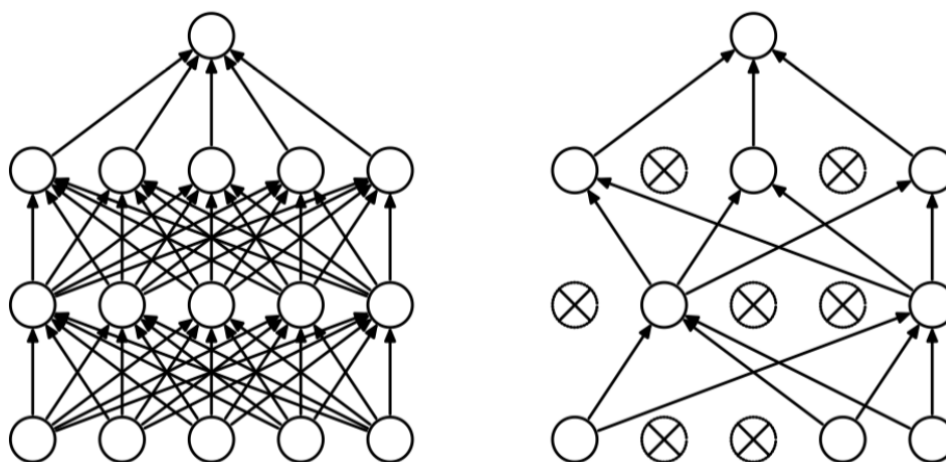
Como já mencionamos, existem algumas técnicas abordadas para a redução do *overfitting*, onde a maioria dessas estratégias para aumento da capacidade da generalização se concentram na própria arquitetura do modelo. Soluções como *dropout* e *batch normalization* foram desenvolvidas para tentar estender o aprendizado profundo para aplicações em conjuntos de dados menores.

A técnica de *dropout* (GOODFELLOW; BENGIO; COURVILLE, 2017) consiste em remover, aleatoriamente a cada iteração de treinamento, uma determinada porcentagem dos neurônios de uma camada, readicionando-os na iteração seguinte. Essa técnica também confere à rede a habilidade de aprender atributos mais robustos, uma vez que um neurônio não pode depender da presença específica de outros neurônios.

Basicamente essa técnica ignora alguns neurônios durante o treinamento, ou seja, diferentes neurônios são removidos da rede temporariamente. A partir da Figura 7, podemos ver que durante a fase de treinamento padrão, todos os neurônios estão envolvidos e durante a aplicação do *dropout*, apenas alguns neurônios selecionados estão

envolvidos com o restante 'desligado'. Assim, após cada iteração, diferentes conjuntos de neurônios são ativados, para evitar que alguns neurônios dominem o processo. Isso, portanto, nos ajuda a reduzir a ameaça de *overfitting* e permite o surgimento de arquiteturas de rede maiores e mais profundas que podem fazer boas previsões sobre dados que a rede não viu antes (HINTON et al., 2012).

Figura 7 – Rede Neural padrão (à esquerda) e Rede Neural aplicando *dropout* (à direita).



A técnica conhecida como Normalização em Lote (em inglês, *Batch Normalization*) foi desenvolvida por Ioffe e Szegedy (2015) para o tratamento de problemas de inicialização da rede. Um dos fatores que dificultam o treinamento é que os valores das funções de ativações das camadas anteriores estão sempre mudando. Neste sentido, a Normalização em Lote normaliza a saída de uma camada de ativação para seguir uma distribuição Gaussiana normal, e como consequência, acelerando o processo de aprendizado da rede. Em algumas circunstâncias a utilização da técnica de *batch normalization* despreza a necessidade da utilização da técnica de *dropout* nas redes convolucionais, diminuindo o tempo de treinamento da rede (LEE; GALLAGHER; TU, 2016).

3.3.1 Hiper-parâmetros

Além dos parâmetros, existem os hiper-parâmetros da CNN. Diferente dos parâmetros que serão ajustados com o treinamento, os hiper-parâmetros são algumas variáveis utilizadas na etapa de configuração da CNN. Os mais comuns são: função de ativação, funções de custo, otimizadores, taxa de aprendizado e tamanho do *mini batch*.

Funções de Ativação: Normalmente, após as camadas convolucionais ou densas, são utilizadas as funções de ativação. Essas funções realizam transformações não-lineares nos dados, de modo a gerar saídas linearmente separáveis. Uma das funções mais comumente utilizadas com este propósito é a *Rectified Linear units* (ReLU) (DAHL; SAINATH; HINTON, 2013), apresentada na Equação 3.1,

$$f(x) = \max(0, x), \quad (3.1)$$

onde x é a entrada para um neurônio. Esta função de ativação substitui os pixels dos mapas de características com valores negativos por zero. Essas camadas também são responsáveis pela etapa de predição dos vetores obtidos nas camadas densas ao final da arquitetura. As funções mais conhecidas são: *softmax* e *sigmoid*.

Treinamento, função de custo e otimizadores: Diante da estrutura apresentada, assim como as RNAs, as CNNs apresentam parâmetros que serão ajustados durante o treinamento. A quantidade total de parâmetros é calculada a partir da soma de todos os parâmetros de cada camada.

Para medir a qualidade atual do treinamento de uma CNN, existe o custo (*loss*) associado a cada época de treino. O valor do custo está associado a qualidade da predição realizada e representa o quão distante os parâmetros atuais estão da predição ideal. Existem diversas funções (*loss function*) propostas no estado da arte, no entanto, a entropia cruzada (*cross-entropy*) é a mais utilizada. Essa função calcula a soma das entropias cruzadas entre a predição e a classe real considerando um único exemplo.

A partir do custo calculado durante a época atual de treino, a rede neural irá ajustar os parâmetros por meio da retropropagação e de algoritmos de otimização. Dentre eles, o mais utilizado é o Gradiente Descendente (GD). Esse algoritmo calcula de forma iterativa para cada parâmetro da rede neural os valores que minimizam a função de objetivo. Após encontrar os valores, a retropropagação irá reajustar os parâmetros. O GD também é responsável por definir a mudança no passo em direção ao gradiente mediante a taxa de aprendizado.

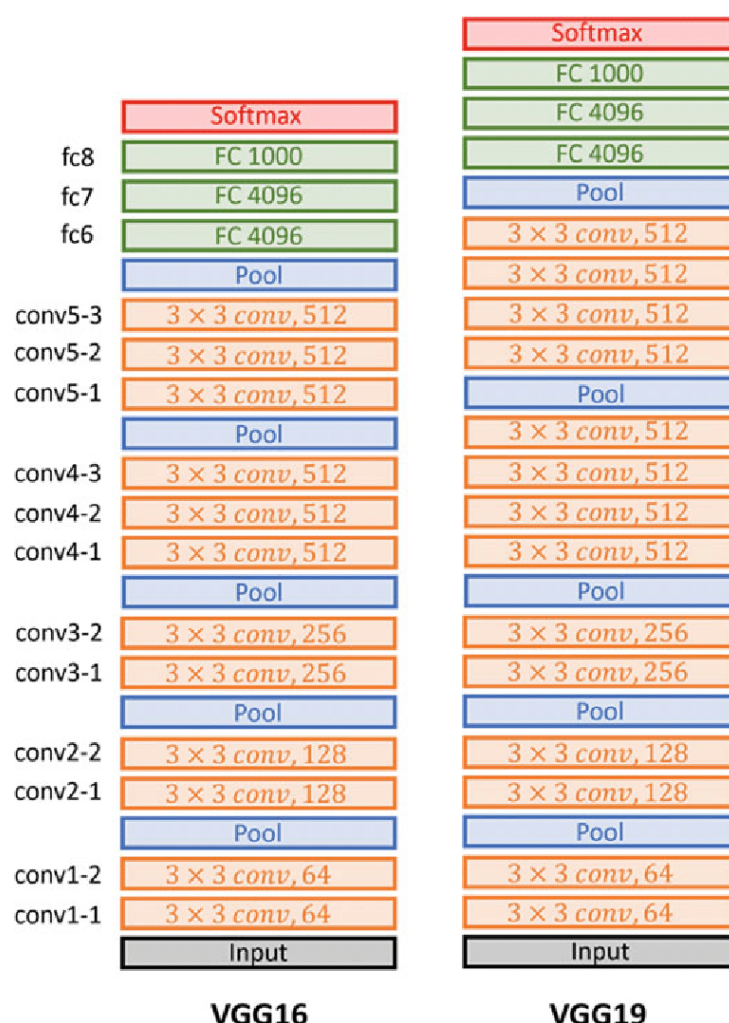
O GD é considerado um algoritmo de otimização simples para CNNs com muitos parâmetros e que necessita de grandes bases de dados para treinamento. Sendo assim, para aplicações mais robustas envolvendo aprendizado profundo, outros algoritmos de otimização como o *Stochastic Gradient Descent* (SGD), Adam, AdaGrad e RMSProp são utilizados. Esses algoritmos possuem diferenças no desempenho e tendem a modificar a taxa de aprendizado com o passar das épocas.

3.3.2 Arquiteturas e ImageNet

A popularidade das CNNs cresceu durante o *ImageNet Large Scale Visual Recognition Challenge* (ILSVRC) (RUSSAKOVSKY et al., 2015). Essa competição ocorreu a partir de 2010 por 7 anos ininterruptos e teve como principais desafios a classificação de imagens e detecção de objetos em larga escala. O *dataset* popularmente conhecido como ImageNet possui mais de um milhão de imagens e cerca de mil classes. Vale destacar que esse evento trata da mais famosa e maior competição em visão

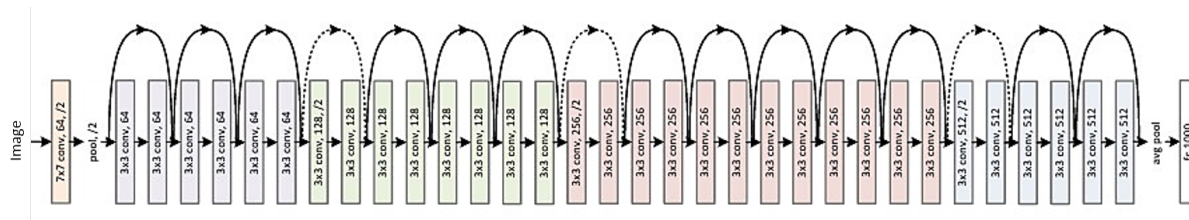
dobra em quase todos os blocos convolucionais, onde apenas o último mantém a mesma quantidade. A arquitetura dispõe de duas camadas totalmente conectadas com 4.096 elementos. A Figura 9 apresenta uma ilustração das redes Vgg-16 (à esquerda) e Vgg-19 (à direita).

Figura 9 – Ilustração das arquiteturas pré-treinadas Vgg-16 (Esquerda) e Vgg-19 (Direita).
Fonte: (SIMONYAN; ZISSERMAN, 2014)



A ResNet50 (HE et al., 2016) é uma arquitetura de rede neural convolucional profunda proposta em 2016 para resolver o problema de gradiente que causa a saturação na aprendizagem e, conseqüentemente, diminui o treinamento. A ideia básica é pular blocos de camadas convolucionais usando conexões de atalho para formar *skip connections* chamados blocos residuais. Esses blocos residuais empilhados melhoram significativamente a eficiência de treinamento e resolvem, principalmente, o problema da degradação presente em redes profundas. Na Figura 10 apresentamos uma ilustração da CNN ResNet.

InceptionV3 (SZEGEDY et al., 2016) é uma arquitetura que emergiu como uma nova versão para as arquiteturas *GoogLeNet* e *InceptionV2*. Essa arquitetura reduz a complexidade da CNN em termos de número de operações realizadas usando

Figura 10 – *Residual Network* com 34 camadas treináveis. Fonte: (HE et al., 2016).

módulos de inicialização, que consiste em combinações paralelas de camadas com filtros convolucionais de tamanho 1×1 , 3×3 e 5×5 . Convoluções com filtros maiores são computacionalmente mais dispendiosos. Portanto, os autores propuseram realizar convoluções 1×1 em primeiro lugar, reduzindo a dimensionalidade do mapa de características e, em seguida, executar convoluções com os outros filtros. O uso de módulos de inicialização resulta em uma redução de 28% no número de parâmetros em comparação com as camadas convolucionais tradicionais.

Xception (CHOLLET, 2017) é uma extensão da arquitetura *Inception* que substitui os módulos de inicialização padrão com *depthwise separable convolution*. Em vez de particionar os dados de entrada em vários blocos compactados, ele mapeia as correlações espaciais para cada canal de saída separadamente. Em seguida, realiza uma convolução de 1×1 em profundidade para capturar a correlação entre os canais. Esta operação é essencialmente equivalente a um processo existente conhecido como *depthwise separable convolution*, que consiste em uma convolução de profundidade (uma convolução espacial realizada de forma independente para cada canal) seguida de uma convolução pontual (filtros com tamanho 1×1 entre canais). A *Xception* alcançou resultados superiores quando comparado às versões anteriores, apesar de ter menos camadas e parâmetros. A inclusão das camadas de *depthwise separable convolution* também forneceu maior eficiência no custo computacional, o que é menos dispendioso e mais rápido que a convolução padrão para realizar um número menor de operações.

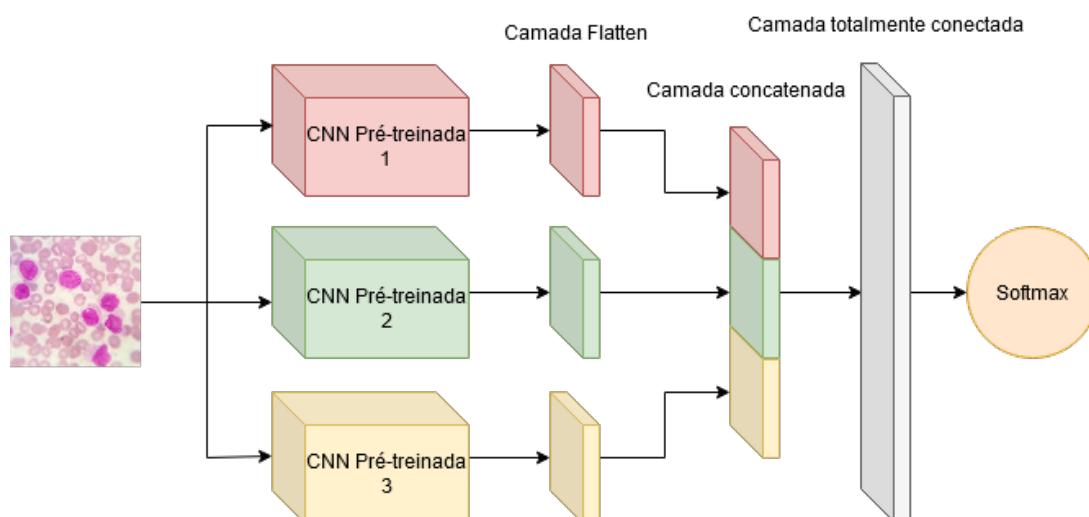
Dense convolutional networks, ou DenseNet, é um bloco denso usado para melhorar o fluxo de informações entre camadas. Esta rede usa menos parâmetros do que a ResNet para seu treinamento (HUANG et al., 2017). Densenet121 consiste em 121 camadas e, nesta arquitetura, cada camada se conecta a todas as camadas subsequentes. Além disso, cada camada recebe características importantes aprendidas por qualquer camada anterior da rede, o que torna o treinamento de rede mais eficiente (LI et al., 2018).

3.4 Multilevel

Devido à variedade de tamanhos e morfologias dos leucócitos, a geração de diversas características de diferentes CNNs são atraentes. Para garantir que não haja a perda de informações úteis, avaliamos neste estudo uma CNN baseada em multilevel, assim, a partir das redes pré-treinadas aplicadas neste estudo é feita a combinação das características de duas ou mais CNNs. A ideia principal por trás da utilização deste conceito é que cada CNN utilizada irá extrair diferentes características. Ao final, essas diversas características são todas concatenadas.

De acordo com [Lyu e Ling \(2018\)](#), todos os mapas de características são concatenados em um vetor unidimensional e, em seguida, conectados a uma camada totalmente conectada, como podemos visualizar na Figura 11. Essa estratégia pode ajudar a melhorar a precisão da classificação dos modelos.

Figura 11 – Esquema genérico do funcionamento do Multilevel.



Essa configuração multilevel é baseada em algumas ideias principais. Primeiro, todos os mapas de características de cada CNN são achatados e concatenados em um vetor unidimensional e, em seguida, conectados a uma camada totalmente conectada. Essa estratégia pode ajudar a melhorar a precisão da classificação do modelo, isso porque cada CNN desempenha um papel importante no resultado final, pois extrai diversas características para representar a imagem de entrada, sejam eles focados apenas nas células ou na totalidade da imagem da lâmina de sangue. Em segundo lugar, esta técnica ajuda a melhorar a taxa de variação dos resultados de classificação. A variância refere-se à sensibilidade do modelo de aprendizado às especificidades dos dados de treinamento, por exemplo, o ruído e as observações específicas. Isso é interessante, pois o modelo será especializado para os dados de entrada aprendendo ruídos aleatórios e será diversificado cada vez que for treinado em dados diferentes ([BELKIN et al., 2019](#)).

3.5 Comitê de Classificadores

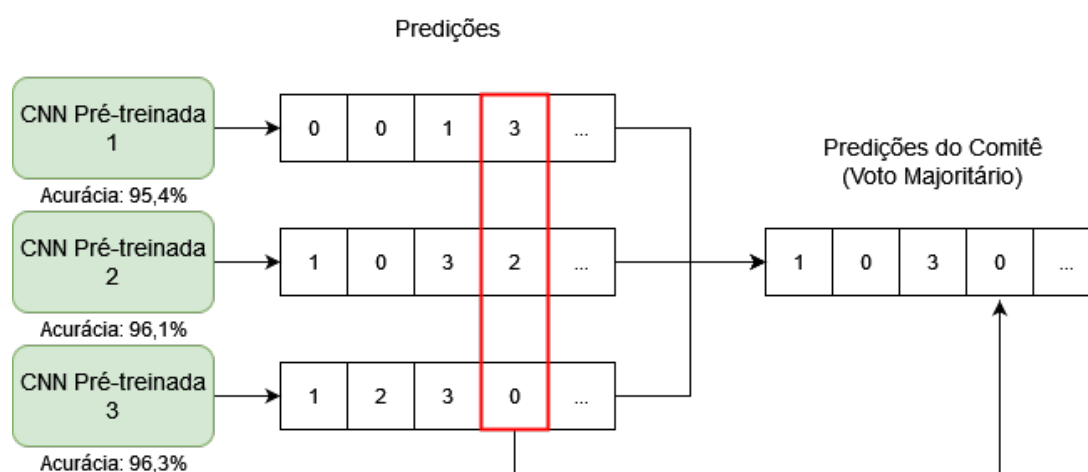
A técnica de comitê foi proposta muito antes do surgimento do paradigma de aprendizado profundo (DASARATHY; SHEELA, 1979). A teoria por trás dessa técnica é bastante simples e apoiada pela conhecida noção de “sabedoria de multidões”: em vez de confiar em apenas um modelo para previsão, é criado um conjunto de vários modelos (pré-treinados). Os resultados dos modelos são então combinados em uma classificação final de seus votos. A ideia original foi desenvolvida para reduzir a variância dos classificadores e obter um melhor desempenho geral (DIETTERICH, 2000).

O voto majoritário é um dos métodos mais populares de classificadores baseados em comitê. Esse método analisa a decisão de cada modelo e seleciona aquela com a maioria dos votos como a decisão final. Ou seja, o rótulo de classe fornecido para uma amostra específica (previsão) será o rótulo de classe que representa a maioria dos rótulos de classes fornecidos para cada classificador (DIETTERICH, 2000).

Como existem problemas multiclases, pode ocorrer um empate ao usar a técnica de voto majoritário. Nesse caso, é utilizado o valor da acurácia das CNNs alcançado durante o estágio de treinamento como um critério de desempate. Por exemplo, se cada CNN no comitê prevê uma classe diferente, a decisão final será a mesma que a CNN que obteve a maior acurácia no estágio de treinamento.

A Figura 12 apresenta como funciona a técnica de comitê adotada, onde existe uma saída de exemplo para quatro elementos de um conjunto de dados. Não houve empate nos três primeiros casos (elemento 1 pertence à classe 1, elemento 2 pertence à classe 0 e elemento 3 pertence à classe 3). No entanto, no caso do quarto elemento, cada CNN votou para uma classe diferente, ocasionando um empate. O resultado final foi a Classe 0, pois foi a saída prevista pela CNN 3, a rede que obteve a melhor acurácia durante o treinamento.

Figura 12 – Exemplo da classificação de comitê adotada.



3.6 Métricas de Avaliação dos Resultados

Para avaliar o desempenho dos algoritmos de classificação foi utilizada a matriz de confusão, onde a partir dela serão calculadas as métricas de avaliação dos resultados.

A matriz de confusão é uma tabela que mostra o resultado da classificação, comparando-o com o resultado real. A matriz, nesse trabalho, é composta por 4 valores: Verdadeiro Positivo (VP), Falso Positivo (FP), Falso Negativo (FN) e Verdadeiro Negativo (VN), onde esses valores podem ser observados na Tabela 3. Tais valores são obtidos a partir do resultado da classificação. Supondo a existência de duas classes hipotéticas: classe X e classe Y, são computados os seguintes valores:

- VP: número de objetos da classe X classificados como sendo da classe X;
- FP: número de objetos da classe Y classificados como X;
- FN: número de objetos da classe X classificados como Y;
- VN: número de objetos da classe Y classificados como sendo da classe Y.

Tabela 3 – Matriz de confusão de duas classes (X,Y).

Valor Previsto \ Valor verdadeiro	Classe X	Classe Y
	VP	FP
Classe X	VP	FP
Classe Y	FN	VN

A partir dessas quantidades, os valores de Precisão (P), *Recall* (R), Acurácia (A), F1-score e Índice Kappa (K) são calculados (POWERS, 2007).

Precisão (P) reflete a proporção de verdadeiros positivos em relação a todas as predições positivas. A precisão de uma classificação mostra a quantidade de objetos da classe X classificados corretamente em relação a todos os objetos classificados como sendo da classe X. Ela pode ser calculada a partir da Equação 3.2,

$$P = \frac{VP}{VP + FP}. \quad (3.2)$$

Recall (R) reflete a proporção de verdadeiros positivos em relação a suas predições positivas e as suas incorretas predições negativas. Essa medida mostra o comportamento dos objetos classificados como X, ou seja, de todos os objetos da classe X, quantos foram classificados como X. Ela pode ser calculada a partir da Equação 3.3,

$$R = \frac{VP}{VP + FN}. \quad (3.3)$$

Acurácia (A) de um classificador é a porcentagem de casos corretamente classificados em um conjunto de teste. Essa medida mede o quão bem um classificador reconhece instâncias de diversas classes. Ela pode ser calculada a partir da Equação 3.4,

$$A = \frac{VP + VN}{VP + FP + VN + FN}. \quad (3.4)$$

F1-score é uma média harmônica entre precisão e *recall*. Com essa informação podemos dizer a performance do classificador com um indicador apenas. Como essa métrica é uma média, esta confere uma visão mais exata da eficiência do classificador do que apenas a precisão ou o *recall*.

$$F1\text{-score} = 2 * \frac{P * R}{P + R}. \quad (3.5)$$

O índice kappa (K) é utilizado como uma medida apropriada da exatidão por representar inteiramente a matriz de confusão. Este índice toma todos os elementos da matriz em consideração, ao invés de apenas aqueles que retratam a quantidade de classificações verdadeiras, o que ocorre quando se calcula a exatidão global da classificação (ROSENFELD; FITZPATRICK-LINS, 1986).

O índice Kappa é um coeficiente de concordância para escalas nominais que mede o relacionamento entre a concordância, além da casualidade e a discordância esperada (ROSENFELD; FITZPATRICK-LINS, 1986). O índice Kappa pode ser calculado com base na Equação 3.6,

$$K = \frac{(\text{observado} - \text{esperado})}{1 - \text{esperado}}. \quad (3.6)$$

Neste caso, “observado” é o valor global para a porcentagem correta, ou seja, o somatório da diagonal principal da matriz dividido pela quantidade de elementos e “esperado” são os valores calculados usando-se os totais de cada linha e cada coluna da matriz de confusão. A qualidade dos resultados podem ser avaliados pelo valor do índice Kappa, observado na Tabela 4 (LANDIS; KOCH, 1977).

Tabela 4 – Nível de exatidão de uma classificação, conforme o valor do índice kappa.

Índice Kappa (K)	Qualidade
$K \leq 0,2$	Ruim
$0,2 < K \leq 0,4$	Razoável
$0,4 < K \leq 0,6$	Bom
$0,6 < K \leq 0,8$	Muito Bom
$K > 0,8$	Excelente

A função de custo foi utilizada neste trabalho como uma informação acerca do

desempenho dos modelos. Esta função é responsável por dizer o quão longe se está da previsão ideal e, portanto, quantifica o “custo” ou “perda” ao aceitar a previsão gerada pelos parâmetros atuais do modelo (JANOCHA; CZARNECKI, 2017).

3.7 Considerações Finais

Neste capítulo foram apresentados os principais materiais e métodos utilizados no desenvolvimento deste trabalho. Foram apresentadas as bases de imagens utilizadas, as técnicas de aumento de dados aplicadas para diminuir o *overfitting*, as CNNs aplicadas nos experimentos, as técnicas de combinação multilevel e comitê e por fim, as métricas de avaliação.

O capítulo seguinte descreve as redes propostas (AlertNet) assim como o método proposto para detecção automática de leucemia e seus tipos em imagens de lâminas de sangue. É mostrado também como os algoritmos explanados nesse capítulo são utilizados para a construção do método.

4 Métodos Propostos

Este capítulo descreve o método proposto neste estudo, que tem por finalidade detectar e avaliar a classificação dos tipos de leucemia em imagens de lâminas de sangue. Devido a gravidade das leucemias agudas, o foco inicial deste estudo foi a detecção dos dois principais tipos dessa leucemia (ALL e AML). Após estes testes iniciais ampliamos o nosso método para a diferenciação entre outros tipos de classificação de leucemia.

Assim, este capítulo está dividido em duas seções, onde em um primeiro momento (Seção 4.1) foram desenvolvidas CNNs (AlertNet e suas variações) com intuito de detectar a leucemia aguda em imagens de lâminas de sangue e logo em sequência formar um comitê de CNNs com o objetivo de melhorar a performance das redes. A avaliação em imagens com leucemia aguda foi o primeiro foco do desenvolvimento deste trabalho devido ao fato da detecção precoce ser de fundamental importância para o tratamento apropriado do câncer.

Após a avaliação do desempenho da rede AlertNet na classificação entre imagens saudáveis e leucemia aguda (ALL e AML), aplicamos essa rede e suas variações para a detecção em outros cenários de classificação de leucemia. Para cada cenário testado realizamos experimentos com e sem aumento de dados (*Data augmentation* - DA) para as variações da AlertNet, bem como para sete redes pré-treinadas disponíveis na literatura. Após estes experimentos, são aplicados os métodos de comitê e multilevel nos nossos experimentos para melhorar o desempenho dos resultados. Assim, a Seção 4.2 apresenta todo o processo para obtenção do método proposto em cada cenário avaliado.

4.1 Método para detecção de leucemia aguda

Com base nas arquiteturas apresentadas na literatura, desenvolvemos um modelo de arquitetura de CNN para diagnosticar os tipos de leucemia aguda com base em imagens de lâminas de sangue. Para elaboração deste modelo proposto foram estudadas as arquiteturas que recentemente obtiveram os melhores resultados na detecção de leucemia, de acordo com os estudos encontrados na literatura. Também foi utilizado o comitê para melhorar o desempenho dos resultados.

Com base nas redes AlexNet (KRIZHEVSKY; SUTSKEVER; HINTON, 2012), CaffeNet (JIA et al., 2014) e VggNet (CHATFIELD et al., 2014), foi desenvolvida a *Acute Leukemias Recognition Network* (AlertNet), uma CNN para a classificação de leucemia em lâminas de sangue.

A AlertNet possui cinco camadas convolucionais sucedidas por camadas de *Batch*

Dessa forma, observa-se que o resíduo gerado tende a propagar as características básicas das imagens durante o treinamento. Trabalhos apresentados na literatura comprovam a eficiência dessa abordagem (HE et al., 2016), (RODRIGUES et al., 2022).

No intuito de desenvolver um modelo que possibilite um treinamento com um menor custo computacional, é proposta a AlertNetX. As camadas do tipo *depth-wise separable convolution* (DSC) foram propostas na arquitetura Xception e proporcionam maior eficiência computacional, uma vez que a quantidade de operações realizadas durante a convolução é menor. O modelo proposto com camadas DSC exige menor tempo de treinamento e menor complexidade.

Observando as vantagens do uso de resíduos e camadas de DSC (propagação das características essenciais da imagem original durante o treinamento e maior eficiência computacional, respectivamente) foi proposta uma CNN com ambas as características. Assim, é adicionada aos testes uma arquitetura CNN chamada AlertNetRX. Essa nova arquitetura envolve a inserção de uma camada residual com camadas de DSC. A Figura 14 apresenta a rede AlertNetRX e ilustra os tipos de camadas.

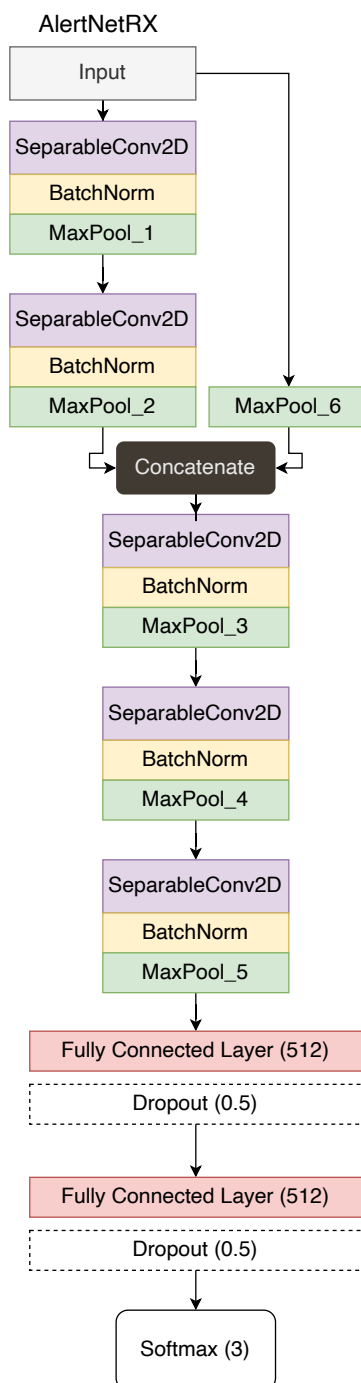
A partir das vantagens da utilização da técnica de comitê, (mencionada na Seção 3.5), foram realizadas algumas combinações de CNNs com esta técnica. A Figura 15 apresenta o comitê que obteve um melhor desempenho dentre as combinações de CNNs testadas. Este comitê consiste em três CNNs: duas pré-treinadas (ResNet50 e InceptionV3) e a rede AlertNetRWD. Após estes experimentos, ampliamos nosso estudo para outras classificações dos tipos de leucemia e a processo de desenvolvimento do método é apresentado a seguir.

4.2 Avaliação do impacto de aumento de dados e uma combinação de CNNs na classificação de leucemia

Devido ao desempenho obtido pela rede AlertNet na classificação de leucemia aguda, foram conduzidos novos testes com esta rede em outros cenários de classificação de tipos de leucemia, e também uma avaliação de algumas técnicas que estão sendo amplamente utilizadas na literatura e o problema que elas abordam. Assim, para uma melhor avaliação e análise dos resultados os testes foram conduzidos em cinco cenários, com base na divisão das classes, a saber:

- Classificação Binária:
 - Cenário 1: leucemia vs. HBS;
 - Cenário 2: ALL vs. HBS;
 - Cenário 3: AML vs. HBS;

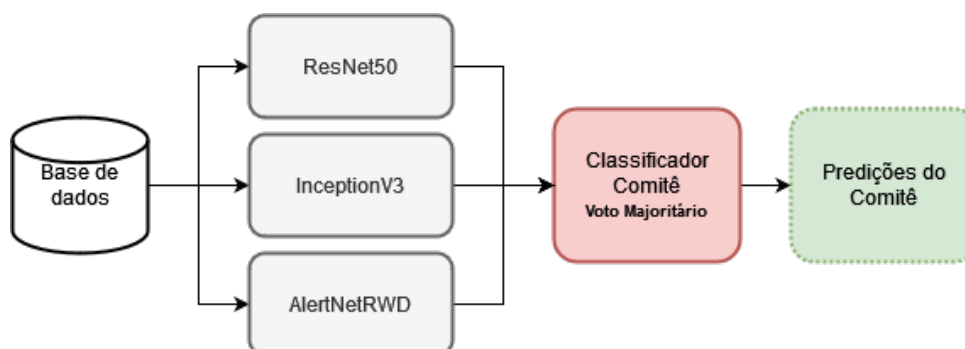
Figura 14 – Topologia do modelo proposto.



- Classificação Multiclasse:
 - Cenário 4: ALL vs. AML vs. HBS;
 - Cenário 5: ALL vs. AML vs. HBS vs. outros tipos.

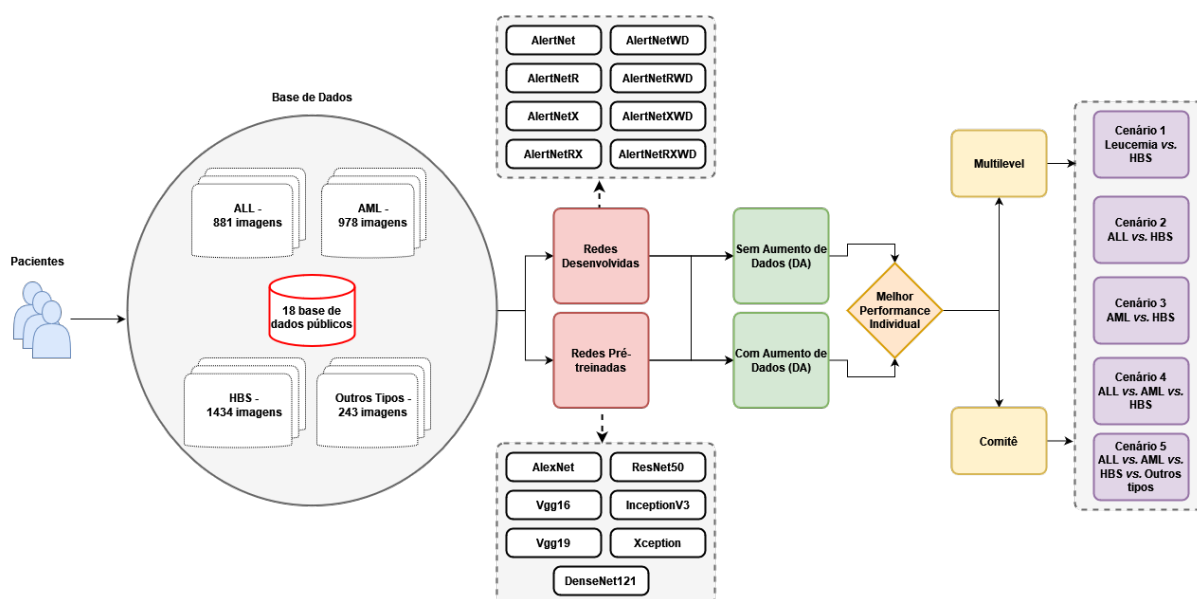
O método de avaliação utilizado neste estudo está de acordo com o fluxograma mostrado na Figura 16. Inicialmente, foi realizada uma pesquisa dos conjuntos de dados disponíveis publicamente, uma vez que a qualidade do conjunto de dados de imagens

Figura 15 – Fluxograma do comitê proposto.



usado tem um impacto significativo no desempenho da classificação. Assim, foram coletadas 3.536 imagens de lâminas de sangue, o que, embora seja um número superior em relação a alguns trabalhos utilizados na literatura, ainda é insuficiente para um treinamento adequado de uma CNN (CHEN, 2019). Portanto, a solução utilizada neste trabalho foi aumentar a capacidade de generalização dos modelos de classificação durante o treinamento baseado na técnica de aumento de dados (*Data Augmentation* - DA). Seguindo o fluxograma do método proposto, realizamos experimentos com as oito redes propostas no nosso estudo e sete redes pré-treinadas que são amplamente utilizadas na literatura. A técnica DA gera novos exemplos para treinamento das CNNs, e o desempenho das redes desenvolvidas e pré-treinadas foram avaliadas com e sem a sua utilização.

Figura 16 – Fluxograma do método de avaliação utilizado para detectar diferentes tipos de leucemia a partir de imagens de lâminas de sangue.



Em seguida, levamos em consideração os cinco cenários em estudo para o conjunto de dados coletados e usamos sete redes neurais para extrair características das imagens de entrada e classificá-los. Depois de realizar todos esses experimentos,

analisamos os resultados individuais de cada rede para cada cenário e aplicamos e avaliamos os métodos multilevel e comitê. Todos os algoritmos apresentados nesta tese foram implementados em Python, com o uso das bibliotecas *scikit-image* (WALT et al., 2014) e *tensorflow* (ABADI et al., 2016).

Neste trabalho foram obtidas dezoito bases de imagens públicas de lâminas de sangue, conforme apresentado na Tabela 2. Devido a divisão em cinco cenários, cada um deles utiliza uma determinada quantidade de imagens. Assim, para o cenário 1 e 5 são utilizadas as 3.536 imagens, já o cenário 2 são utilizadas apenas as imagens saudáveis e com ALL, o que totaliza 2.315 imagens. No cenário 3 são aplicadas 2.412 imagens e no cenário 4 são 3.293 imagens.

4.3 Considerações Finais

O método proposto é composto por um conjunto de ferramentas e procedimentos computacionais. O primeiro experimento do trabalho é focado no desenvolvimento da rede AlertNet para a detecção dos tipos de leucemia aguda. Já o segundo experimento realiza novos testes com esta rede e outras redes pré-treinadas para a avaliação de algumas técnicas que estão sendo aplicadas na detecção dos principais tipos de leucemia. Foram utilizadas dezoito bases de dados públicas com diferentes tipos de leucemia. Redes neurais convolucionais com técnicas de aumento de dados e técnicas como multilevel e comitê foram utilizadas nas imagens com o objetivo de extrair atributos e melhorar o desempenho na classificação.

5 Resultados e Discussões

Este capítulo apresenta os resultados experimentais obtidos em cada etapa do método proposto. O conjunto de dados utilizados no desenvolvimento desta avaliação é composto por imagens que incluem 1.434 imagens de lâminas saudáveis, 881 imagens de ALL, 978 imagens de AML e 243 imagens de outros tipos de leucemia. Aplicamos o *k-fold cross validation* com o valor de k igual a 5 para os resultados obtidos. Todos os experimentos foram realizados em um computador com um processador de 3.6 GHz Intel®Xeon™, 24 GB de RAM, e placa de vídeo NVIDIA TITAN XP 12 GB.

Este capítulo é dividido em duas partes. A primeira parte apresenta os resultados da rede AlertNet e suas variações para a detecção de leucemia aguda. A segunda parte apresenta os resultados das redes testadas para cinco cenários de classificação dos tipos de leucemia.

5.1 Resultados da detecção de leucemia aguda

Os resultados para detecção de leucemia aguda estão estruturados da seguinte forma: primeiro são apresentados os resultados individuais para cada CNN mencionada na Seção 4.1, e em sequência são apresentados os resultados dos modelos que realizam o experimento de comitê de classificadores.

As CNNs ResNet50, InceptionV3 e Xception apresentaram excelentes resultados na competição da ImageNet. Além disso, eles têm um número relativamente baixo de parâmetros quando comparados a outras arquiteturas sequenciais, por isso, foram aplicadas técnicas de ajuste fino (PASUPA; SUNHEM, 2016), que consistem em utilizar uma arquitetura pré-treinada para realizar o aprendizado por transferência. Assim, o treinamento é realizado apenas em camadas selecionadas e com menores taxas de aprendizado.

A técnica de ajuste fino mais utilizada na literatura é o *Shallow fine-tuning* (SFT) (PASUPA; SUNHEM, 2016). No SFT, as camadas iniciais são congeladas, diminuindo a complexidade durante o treinamento da CNN, e apenas as camadas finais são retreinadas. Essas camadas possuem características específicas relacionadas ao conjunto de dados utilizado. Alguns autores afirmam que o SFT não funciona bem quando o domínio alvo difere daquele usado para pré-treinar os pesos (IZADYYAZDANABADI et al., 2018). Por exemplo, imagens fotográficas naturais da ImageNet pertencem a um domínio diferente em comparação com imagens de esfregaço de sangue.

Nessas situações, é melhor aplicar o *Deep Fine-Tuning* (DFT). A abordagem DFT

permite treinar toda a CNN. No entanto, requer tanto um custo computacional, quanto uma quantidade maior de dados. A Tabela 5 apresenta os resultados alcançados, aplicando a técnica SFT e DFT para as redes as ResNet50, InceptionV3 e Xception nas 3.293 imagens (1.434 imagens saudáveis, 881 com ALL e 978 com AML).

Tabela 5 – Resultados obtidos com *Shallow Fine Tuning* e *Deep Fine Tuning*.

Modelo	A (%)	P (%)	R (%)	F1-score	K	Num. param	File size
Shallow Fine Tuning							
ResNet50	43,56±0,03	18,97±0,02	33,33±0	26,43±0,03	0	24.638.339	98mb
InceptionV3	67,66±2,76	68,85±3,41	67,32±3,29	67,82±2,87	0,5081±0,044	22.853.411	92mb
Xception	68,94±3,85	73,38±2,85	64,42±4,13	67,02±4,38	0,5050±0,060	21.912.107	88mb
Deep Fine Tuning							
ResNet50	95,65±1,07	95,81±1,03	95,72±1,05	95,64±1,09	0,9331±0,016	24.638.339	188mb
InceptionV3	94,50±0,76	94,66±0,64	94,52±0,76	94,47±0,79	0,9152±0,011	22.853.411	175mb
Xception	93,56±0,97	93,70±0,91	93,57±0,95	93,50±1,00	0,9006±0,015	21.912.107	167mb

Na Tabela 5, observa-se que o desempenho das CNNs utilizando o SFT foi inferior ao utilizando o DFT. A arquitetura ResNet50, por exemplo, obteve um valor de kappa igual a 0 (zero), o que significa que esta CNN classificou corretamente todas as imagens em apenas uma classe. Comparando os resultados das CNNs utilizando o SFT e o DFT, pode-se perceber que o uso do DFT resultou em um ganho substancial de desempenho das CNNs pré-treinadas. Com DFT, o ResNet50 obtém os melhores resultados em termos de acurácia (95,65%), precisão (95,81%), *recall* (95,72%), F1-score (95,64%) e *kappa* (0,9331).

Na Tabela 6 são apresentados os resultados alcançados pelas arquiteturas propostas neste estudo, a AlertNet e suas variações. Nos resultados, percebe-se que a arquitetura AlertNetRWD alcançou os melhores resultados em termos de acurácia (94,74%), precisão (94,87%), *recall* (93,89%), f1-score (94,72%) e *kappa* (0,9191). Outro ponto a ser destacado desta tabela são os resultados obtidos pelo AlertNet sem *dropout* (WD), onde os resultados são estatisticamente iguais.

Tabela 6 – Resultados obtidos com *K-fold cross validation* para as variações da AlertNet.

Modelo	A (%)	P (%)	R(%)	F1-score(%)	K
AlertNet	93,71±0,68	94,00±0,57	92,62±0,87	93,66±0,70	0,9031±0,010
AlertNetWD	94,56±0,68	94,69±0,66	93,69±0,89	94,53±0,70	0,9163±0,010
AlertNetR	93,68±0,84	93,97±0,71	92,67±1,02	93,63±0,88	0,9027±0,013
AlertNetRWD	94,74±0,86	94,87±0,79	93,89±1,04	94,72±0,87	0,9191±0,013
AlertNetX	92,65±1,14	93,04±1,07	91,64±1,28	92,60±1,14	0,8868±0,017
AlertNetXWD	93,26±1,11	93,39±1,08	92,21±1,31	93,21±1,13	0,8961±0,017
AlertNetRX	90,89±1,15	91,35±1,04	89,80±1,20	90,83±1,17	0,8597±0,017
AlertNetRXWD	92,16±0,73	92,57±0,44	91,12±1,00	92,12±0,76	0,8793±0,011

A partir dos resultados da Tabela 6 é possível inferir que a remoção da camada *dropout* ajuda a obter melhores resultados. O *dropout* é uma técnica de regularização, e seu uso reduz a capacidade de generalização do modelo. Seria necessário aumentar o

tamanho do modelo porque, normalmente, o erro do conjunto de dados de validação é muito menor ao usar *dropout*, mas modelos maiores custam mais iterações de treinamento. Quando o conjunto de dados de treinamento é pequeno, o uso do *dropout* se torna menos eficaz.

Para o desenvolvimento do comitê de CNNs, foi feita uma avaliação das CNNs da literatura e as apresentadas na Seção 4.1. As previsões dos modelos das CNNs pré-treinadas são calculadas para construir um modelo de comitê de classificadores. Uma vantagem de usar esse conjunto de modelos é que ele não precisa de treinamento, e outra vantagem, segundo Géron (2022), é que o comitê geralmente consegue performar melhor do que os estimadores individuais.

Primeiramente, foram avaliadas as variações da arquitetura AlertNet, uma vez que foram propostas para identificar os tipos de leucemia aguda. A Tabela 7 apresenta os resultados alcançados por meio desses comitês. O primeiro comitê foi realizado com as quatro CNNs com a camada de *dropout* (AlertNet, AlertNetR, AlertNetX e AlertNetRX). As CNNs sem a camada de *dropout* (WD) formaram o segundo comitê. Por fim, o terceiro comitê é composto pelas oito arquiteturas AlertNet. Assim, é possível inferir que o comitê utilizando apenas CNNs sem a camada de *dropout* obteve um melhor desempenho.

Tabela 7 – Resultados alcançados com diferentes comitês formados por variações da rede AlertNet.

Modelo Comitê	A (%)	P (%)	R(%)	F1-score(%)	K
AlertNet, AlertNetR, AlertNetX, AlertNetRX	93,31±0,95	93,90±0,84	93,30±0,94	93,33±0,96	0,8971±0,014
AlertNetWD, AlertNetRWD, AlertNetXWD, AlertNetRXWD	94,50±0,83	94,82±0,63	94,52±0,84	94,51±0,82	0,9153±0,013
AlertNet, AlertNetWD, AlertNetR, AlertNetRWD, AlertNetX, AlertNetXWD, AlertNetRX, AlertNetRXWD	94,22±0,73	94,83±0,46	94,23±0,71	94,26±0,71	0,9111±0,011

A Tabela 8 apresenta os resultados da predição dos modelos de comitê com CNNs pré-treinadas. Na primeira linha, o comitê foi formado pelas refinadas CNNs ResNet50, InceptionV3 e Xception. Escolhemos essas CNNs porque obtiveram o melhor desempenho de classificação usando o *Deep Fine Tuning*.

Para melhorar o comitê, foram realizados testes, incluindo as oito versões do AlertNet separadamente. Esses testes visam selecionar uma das oito versões do AlertNet que obtiveram o melhor desempenho para fazer parte do comitê. A partir da análise realizada, verifica-se que a CNN AlertNetRWD obteve um melhor desempenho. Por fim, observa-se que o melhor conjunto de CNNs para compor o comitê é mostrado na segunda linha da Tabela 8. Essa é a união das redes pré-treinadas ResNet50, InceptionV3 e AlertNetRWD.

Tabela 8 – Resultados alcançados com diferentes comitês formados por CNNs pré-treinadas.

Modelo Comitê	A (%)	P (%)	R (%)	F1-score (%)	K
ResNet50, InceptionV3, Xception	95,38±0,88	95,71±0,59	95,39±0,87	95,40±0,83	0,9288±0,013
AlertNetRWD, ResNet50, InceptionV3	96,17±0,56	96,38±0,43	95,94±0,55	96,18±0,54	0,9411±0,008

5.1.1 Discussões

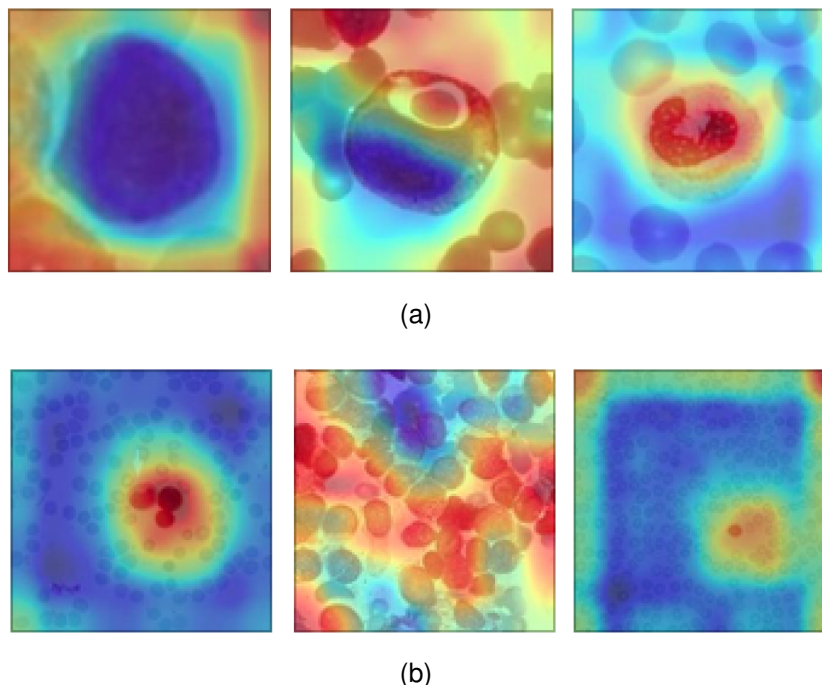
Com base nos resultados obtidos, foi feita uma comparação entre o desempenho do comitê proposto e os desempenhos obtidos por trabalhos da literatura que abordam o mesmo problema. A partir da Tabela 9, pode-se perceber que o estado da arte apresentou valores de acurácia superiores à abordagem proposta. No entanto, existem trabalhos que possuem imagens segmentadas e centralizadas em apenas uma célula, como o trabalho de Ghaderzadeh et al. (2022), o que dificulta o aprendizado da rede para outras entradas, pois faz com que ela aprenda a focar apenas no centro da imagem, então caso venha uma imagem com várias células e em diferentes locais, a rede fica confusa a qual classe pertence a imagem. Outro ponto essencial a ser destacado está relacionado à variabilidade das imagens, pois são provenientes de dezoito conjuntos de dados diferentes. Essa característica leva a uma maior diversidade nos dados de treinamento, o que leva à obtenção de um método robusto para diferentes tipos de imagem de entrada, pois faz com que as redes aprendam a não focar apenas em uma determinada região da imagem. Um exemplo do foco da rede AlertNetRWD para as três classes (ALL-AML-HBS), pode ser observada na Figura 17 através dos mapas de ativação. É possível identificar quais regiões diferenciam imagens saudáveis daquelas com leucemia aguda (linfóide ou mieloide).

Tabela 9 – Comparação entre os resultados obtidos pelo método proposto com os obtidos por métodos relacionados.

Método	Descritores	N. de imagens	Acurácia
Rawat et al. (2017)	Geometria, cor e textura	420	99,50%
Laosai e Chamnongthai (2018)	Forma, cor, Textura e número de nucleoli	500	99,85%
Tran et al. (2018)	CNN	1.636	97,30%
Karar, Alotaibi e Alotaibi (2022)	CNN	445	95,50%
Abhishek et al. (2022)	CNN	608	98,00%
Ghaderzadeh et al. (2022)	CNN	12.528	98,50%
Método proposto	CNN	3.293	96,17%

O número de leucócitos pode variar dependendo da imagem de entrada. A Figura 17 mostra que a AlertNetRWD gera diferentes padrões de mapa de ativação para cada uma dessas situações. Além disso, como é treinado em diferentes conjuntos de

Figura 17 – Exemplos de mapas de ativação para lâminas de sangue, (a) imagens com um leucócito, (b) imagens com vários leucócitos. A primeira coluna contém imagens da classe ALL; a segunda coluna representa a classe AML e a terceira a classe HBS.



dados, o modelo proposto pode se adaptar a diferentes características; é possível observar que os mapas diferem de um conjunto de dados para outro. No entanto, a CNN ativa regiões diferentes para cada classe. Determinadas regiões são consideradas de baixa importância para a classificação final. Por exemplo, na Figura 17(a), da classe ALL, o leucócito ficou predominantemente azul, onde essa cor informa que foi a região mais crítica/importante para a classificação final. Esse padrão muda nas demais classes, tornando-se predominantemente vermelho e de maior importância na classe HBS.

5.2 Resultados da avaliação do impacto do aumento de dados e uma combinação de CNNs na classificação de leucemia

A partir dos resultados obtidos no primeiro experimento, decidiu-se investigar, em um primeiro momento, a influência do uso de aumento de dados (*Data Augmentation* - (DA)) no treinamento das redes para a classificação em cinco cenários de tipos de leucemia, por isso foi feita uma comparação da AlertNet e suas variações, bem como os resultados obtidos pelas sete redes mencionadas na Seção 3.3 com a utilização de DA e sem a utilização de DA.

5.2.1 Classificação Binária

Os experimentos para a classificação binária são subdivididos em três subproblemas: Leucemia vs. HBS; ALL vs. HBS; AML vs. HBS. Esses subproblemas foram escolhidos e analisados devido a dois fatores: o primeiro fator corresponde aos estudos encontrados na literatura que realizam a mesma classificação; já o segundo fator corresponde a existência de conjuntos de dados públicos para cada classe. As subseções seguintes abordam os resultados para estes três cenários.

5.2.1.1 Cenário 1: Leucemia vs. HBS

Para este cenário, foram utilizados 18 conjuntos de dados, totalizando 3.536 imagens, sendo 1.434 da classe HBS e 2.102 da classe Leucemia. A Tabela 10 apresenta os resultados para as sete redes neurais em dois contextos: com e sem aumento de dados (DA). Nesse cenário, a rede ResNet50-DA foi a que obteve o melhor resultado para todas as métricas. Também pode-se observar que a rede AlexNet obteve um bom resultado, contudo seus resultados são inferiores quando comparados às demais redes.

Tabela 10 – Resultados obtidos para leucemia vs. HBS com e sem aumento de dados aplicando *K-fold cross validation*.

Modelo	A(%)	loss	P(%)	R(%)	F1-score(%)	K
AlexNet	95,87±0,51	0,1295	96,09±1,12	96,99±1,34	96,53±0,44	0,9141±0,010
AlexNet-DA	95,25±0,81	0,1296	95,38±1,41	96,70±2,15	96,01±0,71	0,9011±0,016
Vgg16	99,15±0,22	0,0380	99,00±0,30	99,56±0,35	99,28±0,18	0,9823±0,004
Vgg16-DA	99,23±0,29	0,0332	99,04±0,28	99,66±0,27	99,35±0,24	0,9841±0,006
Vgg19	98,98±0,36	0,0404	99,09±0,31	99,18±0,40	99,13±0,31	0,9788±0,007
Vgg19-DA	99,23±0,15	0,0319	99,18±0,12	99,51±0,28	99,35±0,13	0,9841±0,003
ResNet50	99,43±0,26	0,0261	99,37±0,36	99,66±0,13	99,52±0,22	0,9882±0,005
ResNet50-DA	99,54±0,27	0,0220	99,47±0,53	99,75±0,23	99,61±0,23	0,9904±0,005
InceptionV3	98,58±0,36	0,0486	98,89±0,54	98,71±0,31	98,80±0,30	0,9706±0,007
InceptionV3-DA	99,23±0,15	0,0327	99,28±0,35	99,40±0,13	99,34±0,13	0,9838±0,003
Xception	98,19±0,68	0,0681	98,19±0,63	98,75±0,51	98,47±0,57	0,9623±0,014
Xception-DA	98,16±0,51	0,0695	98,56±0,66	98,32±0,47	98,44±0,42	0,9618±0,010
DenseNet121	99,32±0,27	0,0283	99,28±0,16	99,56±0,39	99,42±0,23	0,9858±0,005
DenseNet121-DA	99,37±0,16	0,0320	99,28±0,28	99,66±0,27	99,47±0,13	0,9870±0,003

A Tabela 11 apresenta os resultados obtidos com a rede AlertNet e suas variações para o mesmo conjunto de dados aplicando o DA no treinamento das redes. As redes AlertNetWD-DA e AlertNetRWD-DA obtiveram os melhores desempenhos dentre as redes testadas, com 98,84% e 98,78% de acurácia, respectivamente.

5.2.1.2 Cenário 2: ALL vs. HBS

Existe uma boa variedade de artigos na literatura abordando esse cenário, visto que existem muitos conjuntos de dados públicos disponíveis para a leucemia linfóide aguda. Assim, foram utilizados 14 conjuntos de dados com 2.315 imagens, sendo 1.434

Tabela 11 – Resultados obtidos pela variações da AlertNet para o cenário leucemia vs. HBS, com e sem aumento de dados aplicando *k-fold cross validation*.

Modelo	A(%)	Loss	P(%)	R(%)	F1-score(%)	K
AlertNet	97,96±0,40	0,0718	97,69±0,78	97,27±0,79	97,48±0,50	0,9577±0,008
AlertNet-DA	98,16±0,34	0,0603	98,07±0,72	97,38±0,54	97,72±0,42	0,9617±0,007
AlertNetWD	98,16±0,34	0,0981	98,07±0,72	97,38±0,54	97,72±0,42	0,9617±0,007
AlertNetWD-DA	98,84±0,44	0,0344	98,40±0,90	98,74±0,31	98,57±0,53	0,9759±0,009
AlertNetR	97,79±0,45	0,0829	97,22±0,97	97,34±0,87	97,28±0,55	0,9541±0,009
AlertNetR-DA	98,50±0,29	0,0493	98,32±0,73	97,97±0,90	98,14±0,36	0,9688±0,006
AlertNetRWD	96,94±0,34	0,1045	95,53±0,47	97,00±0,53	96,25±0,41	0,9367±0,007
AlertNetRWD-DA	98,78±0,40	0,0408	98,60±0,54	98,39±0,94	98,49±0,50	0,9747±0,008
AlertNetX	97,48±0,32	0,0847	96,73±0,69	97,06±0,63	96,90±0,39	0,9477±0,006
AlertNetX-DA	97,68±0,23	0,0818	97,28±1,04	96,99±0,83	97,13±0,28	0,9518±0,004
AlertNetXWD	96,86±0,70	0,0965	96,16±1,05	96,09±1,05	96,12±0,86	0,9348±0,014
AlertNetXWD-DA	98,19±0,39	0,0624	98,17±0,44	97,34±1,03	97,75±0,49	0,9623±0,008
AlertNetRX	96,66±0,65	0,1076	95,10±1,89	96,79±0,99	95,92±0,75	0,9309±0,013
AlertNetRX-DA	96,63±0,50	0,0995	95,38±0,82	96,37±0,93	95,87±0,61	0,9302±0,001
AlertNetRXWD	96,31±1,06	0,1108	95,40±1,28	95,53±1,50	95,46±1,31	0,9236±0,022
AlertNetRXWD-DA	97,17±0,81	0,0842	97,50±0,51	95,46±1,72	96,46±1,04	0,9410±0,017

saudáveis e 881 com ALL. A Tabela 12 indica o desempenho obtido pelas sete redes pré-treinadas com e sem DA. Os dados apresentados nesta tabela mostram que a rede DenseNet121-DA obteve o melhor resultado em termos de acurácia, *Recall*, *F1-score* e *Kappa*. Por outro lado, a rede ResNet50-DA obteve a melhor precisão com um valor de 99,79%, contudo os resultados de ambas as redes são estatisticamente semelhantes

Tabela 12 – Resultados obtidos para ALL vs. HBS com e sem aumento de dados aplicando *K-fold cross validation*.

Modelo	A(%)	loss	P(%)	R(%)	F1-score(%)	K
AlexNet	96,19±0,69	0,1058	96,74±0,94	97,14±0,96	96,93±0,55	0,9192±0,014
AlexNet-DA	97,32±0,69	0,0789	97,57±1,10	97,81±1,42	97,68±0,56	0,9392±0,014
Vgg16	99,35±0,43	0,0326	99,44±0,52	99,51±0,31	99,47±0,34	0,9862±0,009
Vgg16-DA	99,52±0,18	0,0758	99,71±0,28	99,50±0,30	99,61±0,14	0,9898±0,003
Vgg19	99,08±0,21	0,0365	99,30±0,35	99,30±0,35	99,29±0,17	0,9816±0,004
Vgg19-DA	99,57±0,30	0,0255	99,78±0,19	99,51±0,39	99,65±0,24	0,9908±0,006
ResNet50	99,61±0,28	0,0204	99,65±0,24	99,71±0,29	99,68±0,22	0,9917±0,005
ResNet50-DA	99,74±0,38	0,0114	99,79±0,31	99,79±0,31	99,79±0,31	0,9944±0,008
InceptionV3	99,30±0,41	0,0310	99,23±0,56	99,64±0,24	99,40±0,33	0,9852±0,008
InceptionV3-DA	99,43±0,12	0,0261	99,51±0,39	99,57±0,38	99,54±0,09	0,9880±0,002
Xception	98,79±0,58	0,0465	98,54±0,56	99,50±0,39	99,02±0,47	0,9742±0,012
Xception-DA	98,70±0,48	0,0493	98,61±0,53	99,30±0,35	98,95±0,38	0,9696±0,011
DenseNet121	99,69±0,24	0,0205	99,65±0,42	99,85±0,18	99,75±0,19	0,9935±0,005
DenseNet121-DA	99,78±0,26	0,0117	99,78±0,31	99,85±0,18	99,82±0,21	0,9954±0,005

A Tabela 13 apresenta os resultados da rede AlertNet e suas variações para o cenário ALL vs. HBS. O melhor desempenho foi da rede AlertNetRWD-DA que obteve os melhores resultados para as taxas de acurácia, recall, f1-score e kappa. A rede AlertNetXWD-DA obteve o melhor resultado na métrica de precisão com 99,31%.

Tabela 13 – Resultados obtidos pelas variações da AlertNet para o cenário ALL vs. HBS com e sem aumento de dados aplicando *k-fold cross-validation*.

Modelo	A(%)	loss	P(%)	R(%)	F1-score(%)	K
AlertNet	98,35±0,39	0,0652	98,28±0,55	97,38±1,30	97,82±0,53	0,9650±0,008
AlertNet-DA	98,87±0,55	0,0534	98,64±1,16	98,40±1,09	98,51±0,72	0,9761±0,011
AlertNetWD	97,36±0,77	0,0877	97,44±0,66	95,56±1,81	96,49±1,05	0,9437±0,016
AlertNetWD-DA	99,35±0,30	0,0250	99,20±0,93	99,08±0,76	99,14±0,39	0,9862±0,006
AlertNetR	97,97±0,39	0,0760	97,60±0,73	97,04±0,73	97,32±0,51	0,9568±0,008
AlertNetR-DA	99,09±0,51	0,0345	98,85±0,56	98,74±1,01	98,80±0,68	0,9807±0,010
AlertNetRWD	96,76±0,34	0,1056	97,97±0,64	93,40±1,30	95,63±0,49	0,9305±0,007
AlertNetRWD-DA	99,39±0,31	0,0298	98,99±1,06	99,42±0,69	99,20±0,41	0,9871±0,006
AlertNetX	98,14±0,99	0,0674	98,26±1,16	98,81±1,73	97,53±1,33	0,9604±0,021
AlertNetX-DA	98,61±0,58	0,0535	98,73±0,47	97,60±1,46	98,16±0,77	0,9705±0,012
AlertNetXWD	96,71±0,67	0,0901	97,73±0,66	93,51±1,17	95,58±0,91	0,9296±0,014
AlertNetXWD-DA	98,96±0,61	0,0422	99,31±1,00	97,94±1,68	98,62±0,83	0,9779±0,013
AlertNetRX	97,92±0,72	0,0659	98,71±0,75	95,79±1,53	97,22±0,98	0,9556±0,015
AlertNetRX-DA	97,88±0,59	0,0691	97,17±1,22	97,27±0,93	97,21±0,78	0,9550±0,012
AlertNetRXWD	96,80±1,21	0,0962	96,69±1,99	94,73±3,05	95,68±1,86	0,9308±0,029
AlertNetRXWD-DA	98,61±0,32	0,0466	98,51±0,83	97,83±1,01	98,17±0,43	0,9705±0,006

5.2.1.3 Cenário 3: AML vs. HBS

Para o último cenário de classificação binária, foram utilizados 16 conjuntos de dados com um total de 2.412 imagens, sendo 1.434 saudáveis e 978 com AML. A Tabela 14 apresenta o desempenho obtido pelas sete redes com e sem DA. Os dados apresentados nesta tabela mostram que a rede ResNet50-DA obteve o melhor resultado em termos de acurácia, precisão, F1-score e Kappa. Por outro lado, as redes DenseNet121-DA e Vgg16-DA apresentaram o melhor valor de *recall*, que foi da ordem de 99,71%.

Tabela 14 – Resultados obtidos para AML vs. HBS com e sem aumento de dados aplicando *K-fold cross validation*.

Modelo	A(%)	loss	P(%)	R(%)	F1-score(%)	K
AlexNet	96,81±0,47	0,1141	97,48±0,78	97,13±1,08	97,30±0,41	0,9338±0,009
AlexNet-DA	96,93±0,71	0,0965	96,78±0,97	98,11±1,53	97,43±0,60	0,9362±0,014
Vgg16	99,17±0,29	0,0354	99,23±0,28	99,50±0,39	99,30±0,24	0,9827±0,006
Vgg16-DA	99,66±0,18	0,0191	99,71±0,15	99,71±0,28	99,71±0,15	0,9931±0,003
Vgg19	99,04±0,45	0,0391	99,09±0,39	99,29±0,65	99,19±0,38	0,9802±0,009
Vgg19-DA	99,42±0,22	0,0271	99,64±0,24	99,36±0,51	99,50±0,19	0,9879±0,004
ResNet50	99,46±0,47	0,0274	99,64±0,34	99,43±0,52	99,54±0,40	0,9888±0,009
ResNet50-DA	99,70±0,18	0,0217	99,92±0,15	99,57±0,38	99,75±0,15	0,9939±0,003
InceptionV3	98,59±0,17	0,0513	98,27±0,47	99,30±0,24	98,78±0,17	0,9698±0,004
InceptionV3-DA	99,46±0,23	0,0349	99,43±0,30	99,64±0,24	99,54±0,19	0,9887±0,004
Xception	98,05±0,63	0,0834	97,73±0,81	99,02±0,29	98,37±0,52	0,9594±0,013
Xception-DA	97,80±0,27	0,1902	97,13±0,48	99,23±0,57	98,16±0,23	0,9541±0,005
DenseNet121	99,42±0,30	0,0302	99,64±0,24	99,37±0,45	99,50±0,26	0,9879±0,006
DenseNet121-DA	99,67±0,10	0,0179	99,65±0,15	99,71±0,28	99,71±0,09	0,9931±0,002

A Tabela 15 apresenta o desempenho da AlertNet e suas variações para o cenário AML vs. HBS. Assim como nos outros dois cenários anteriores as redes AlertNetRWD-DA

e AlertNetWD-DA obtiveram os melhores resultados com acurácia de 99,29% e 99,25%, respectivamente.

Tabela 15 – Resultados obtidos para o cenário AML vs. HBS com e sem aumento de dados aplicando *k-fold cross-validation*.

Modelo	A(%)	loss	P(%)	R(%)	F1-score(%)	K
AlertNet	98,47±0,34	0,0596	98,75±0,46	97,43±0,72	98,09±0,43	0,9681±0,007
AlertNet-DA	99,25±0,40	0,0310	98,77±0,57	99,37±0,55	99,07±0,49	0,9845±0,008
AlertNetWD	97,68±0,40	0,0783	98,32±0,75	95,90±0,95	97,09±0,50	0,9516±0,008
AlertNetWD-DA	99,25±0,11	0,0283	99,17±0,45	99,17±0,27	99,07±0,14	0,9845±0,002
AlertNetR	98,54±0,32	0,0649	98,35±0,83	97,94±0,36	98,15±0,42	0,9690±0,007
AlertNetR-DA	99,21±0,32	0,0438	99,07±0,83	98,97±0,36	99,02±0,42	0,9836±0,007
AlertNetRWD	96,97±0,34	0,0438	96,79±0,55	95,70±0,36	96,24±0,42	0,9371±0,007
AlertNetRWD-DA	99,29±0,34	0,0335	98,58±0,94	99,68±0,45	99,13±0,41	0,9854±0,007
AlertNetX	97,97±0,44	0,0519	98,43±0,81	96,52±0,55	97,46±0,55	0,9577±0,009
AlertNetX-DA	98,54±0,52	0,0519	98,26±1,21	98,15±0,58	98,20±0,65	0,9698±0,010
AlertNetXWD	97,51±0,58	0,0834	97,12±1,15	96,82±0,55	96,92±0,71	0,9483±0,012
AlertNetXWD-DA	99,00±0,17	0,0438	98,77±0,45	98,76±0,58	98,76±0,21	0,9793±0,003
AlertNetRX	97,30±0,32	0,0898	97,60±0,58	95,70±0,68	96,64±0,41	0,9439±0,006
AlertNetRX-DA	98,01±0,43	0,0647	97,87±1,71	97,23±1,28	97,53±0,51	0,9586±0,008
AlertNetRXWD	97,34±0,59	0,0958	97,80±0,64	95,59±1,71	96,68±0,77	0,9447±0,012
AlertNetRXWD-DA	98,34±0,38	0,0578	98,45±0,37	97,74±0,93	98,09±0,60	0,9681±0,009

Pelos resultados obtidos para os três cenários de classificação binária, podemos perceber uma melhora no desempenho das métricas com a aplicação das técnicas de DA em praticamente todas as redes pré-treinadas avaliadas. Realizando um comparativo das redes AlertNet e suas variações com e sem DA, podemos observar um perceptível aumento nas taxas com a utilização de técnicas de aumento de dados. Comparando os resultados das redes pré-treinadas com as redes propostas neste estudo conseguimos confirmar que os resultados são estatisticamente similares.

5.2.2 Classificação Multiclasse

Para classificação multiclasse, foram estudados dois cenários. O primeiro cenário foi dividido em três classes (ALL vs. AML vs. HBS) e o segundo com quatro classes (ALL vs. AML vs. HBS vs. outros tipos). Neste último cenário foi especificada a classe “outros tipos” devido ao pequeno número de imagens disponíveis para diferentes tipos de leucemia, como a leucemia crônica. Assim, se fosse dividida ainda mais essa classe, ela se tornaria ainda mais desbalanceada. Até o momento não foi encontrado nenhum trabalho que realizasse esta divisão, portanto, foram realizados mais testes para este cenário.

5.2.2.1 Cenário 4: ALL vs. AML vs. HBS

Para este cenário, foram utilizadas 3.293 imagens de 18 conjuntos de dados, sendo 1.434 imagens da classe HBS, 881 da classe ALL e 978 da classe AML. Os resultados com as setes redes pré-treinadas obtidos para este cenário são observados

na Tabela 16. Realizando uma comparação com os experimentos que abordam a classificação binária, os resultados obtidos indicaram um desempenho reduzido para esta classificação multiclasse. A rede DenseNet121-DA obteve o melhor desempenho com 97,11% de acurácia.

Também foram realizados testes neste cenário para as redes AlertNet bem como suas variações, e seus resultados podem ser vistos na Tabela 17. O desempenho destas redes diminuiu em relação às redes pré-treinadas, à exceção das redes AlexNet e Xception. A rede AlertNetRWD-DA alcançou 94,74% de acurácia e 94,87% de precisão.

Tabela 16 – Resultados obtidos para ALL vs. AML vs. HBS com e sem aumento de dados aplicando *K-fold cross validation*.

Modelo	A(%)	loss	P(%)	R(%)	F1-score(%)	K
AlexNet	89,58±1,44	0,3499	89,93±1,23	89,58±1,44	89,50±1,51	0,8394±0,022
AlexNet-DA	89,58±1,69	0,2740	90,00±1,38	89,71±1,74	89,42±1,88	0,8389±0,026
Vgg16	96,41±0,69	0,1457	96,49±0,71	96,42±0,69	96,41±0,70	0,9448±0,010
Vgg16-DA	96,99±0,82	0,1083	97,01±0,84	97,00±0,82	96,99±0,81	0,9537±0,012
Vgg19	96,05±0,55	0,1835	96,09±0,55	96,05±0,54	96,04±0,56	0,9392±0,008
Vgg19-DA	96,81±0,56	0,1270	96,86±0,59	96,82±0,58	96,80±0,57	0,9509±0,008
ResNet50	95,65±1,07	0,1508	95,81±1,03	95,72±1,05	95,64±1,09	0,9331±0,016
ResNet50-DA	96,69±0,46	0,1191	96,70±0,44	96,69±0,45	96,68±0,45	0,9490±0,007
InceptionV3	94,50±0,76	0,2061	94,66±0,64	94,52±0,76	94,47±0,79	0,9152±0,011
InceptionV3-DA	96,51±0,56	0,1294	96,58±0,49	96,52±0,53	96,49±0,58	0,9462±0,008
Xception	93,56±0,97	0,1835	93,70±0,91	93,57±0,95	93,50±1,00	0,9006±0,015
Xception-DA	93,50±1,52	0,1687	93,65±1,40	93,55±1,45	93,45±1,56	0,8997±0,023
DenseNet121	96,05±0,91	0,1412	96,15±0,89	96,09±0,89	96,04±0,93	0,9434±0,006
DenseNet121-DA	97,11±0,53	0,0911	97,13±0,55	97,12±0,55	97,11±0,54	0,9556±0,008

Tabela 17 – Resultados obtidos para o cenário ALL vs. AML vs. HBS com e sem aumento de dados aplicando *k-fold cross-validation*.

Modelo	A(%)	loss	P(%)	R(%)	F1-score(%)	K
AlertNet	93,68±0,51	0,2273	93,99±0,51	93,74±0,58	93,61±0,55	0,9025±0,007
AlertNet-DA	93,71±0,68	0,1696	94,00±0,57	92,62±0,87	93,66±0,70	0,9031±0,010
AlertNetWD	91,97±0,80	0,2733	92,23±0,72	91,71±0,86	91,93±0,87	0,8766±0,013
AlertNetWD-DA	94,56±0,68	0,1436	94,69±0,66	93,69±0,89	94,53±0,70	0,9163±0,010
AlertNetR	93,07±0,98	0,2592	93,47±0,87	92,95±0,54	93,04±1,00	0,8933±0,015
AlertNetR-DA	93,68±0,84	0,1665	93,97±0,71	92,67±1,02	93,63±0,88	0,9027±0,013
AlertNetRWD	91,15±1,18	0,3387	91,44±1,19	91,11±1,14	91,08±1,22	0,8634±0,018
AlertNetRWD-DA	94,74±0,86	0,1330	94,87±0,79	93,89±1,04	94,72±0,87	0,9191±0,013
AlertNetX	92,62±0,86	0,2407	92,91±0,83	92,70±0,84	92,57±0,88	0,8863±0,013
AlertNetX-DA	92,65±1,14	0,1985	93,04±1,07	91,64±1,28	92,60±1,14	0,8868±0,017
AlertNetXWD	91,86±1,20	0,2692	92,08±1,17	91,82±1,19	91,82±1,22	0,8745±0,018
AlertNetXWD-DA	93,26±1,11	0,1750	93,39±1,08	92,21±1,31	93,21±1,13	0,8961±0,017
AlertNetRX	91,43±1,04	0,2547	91,77±1,09	91,39±1,02	91,35±1,02	0,8679±0,015
AlertNetRX-DA	90,89±1,15	0,2003	91,35±1,04	89,80±1,20	90,83±1,17	0,8597±0,017
AlertNetRXWD	91,04±1,80	0,3288	91,42±1,47	91,13±1,70	90,99±1,75	0,8619±0,027
AlertNetRXWD-DA	92,16±0,73	0,1934	92,57±0,44	91,12±1,00	92,12±0,76	0,8793±0,011

5.2.2.2 Cenário 5: ALL vs. AML vs. HBS vs. outros tipos

O conjunto de dados utilizado no desenvolvimento deste cenário é composto por imagens que incluem 1.434 imagens de lâminas saudáveis, 881 imagens de ALL, 978

imagens de AML e 243 imagens de outros tipos de leucemia. Assim como nos outros cenários, também é avaliado o uso de aumento de dados. A Tabela 18 apresenta o resultado desta avaliação, onde pode-se observar uma melhoria em termos de métricas de desempenho em todas as redes avaliadas com o uso de DA. A arquitetura DenseNet121-DA foi a que obteve o melhor desempenho para a acurácia, precisão, *recall* e F1-score obtendo uma taxa igual ou superior a 94%.

Cada aumento de dados pode ter um impacto diferente em termos de robustez geral que ele traz para uma rede neural. É de conhecimento comum que algumas técnicas de aumento de dados melhoram a estabilidade de decisão da rede neural utilizada, mas outras podem ter um efeito prejudicial (SHORTEN; KHOSHGOFTAA, 2019). Com base nisso e nos resultados obtidos, foram realizados testes individuais para cada técnica de DA com as sete arquiteturas de rede. Cada rede foi treinada em cinco cenários, conforme são aplicadas as cinco técnicas de aumento de dados: rotação, translação, espelhamento, zoom e cisalhamento.

Tabela 18 – Resultados obtidos para ALL vs. AML vs. HBS vs. outros tipos com e sem Aumento de Dados aplicando K-fold cross validation.

Modelo	A(%)	loss	P(%)	R(%)	F1-score(%)	K
AlexNet	82,74±0,89	0,5436	82,14±1,16	82,76±0,89	81,46±0,91	0,7461±0,013
AlexNet-DA	84,46±1,05	0,4237	84,29±1,55	84,28±1,16	83,07±1,30	0,7713±0,016
Vgg16	92,72±1,44	0,2129	92,87±1,42	92,73±1,49	92,72±1,41	0,8946±0,020
Vgg16-DA	92,95±0,93	0,2140	92,98±0,74	92,94±0,93	92,88±0,87	0,8979±0,013
Vgg19	91,30±1,02	0,2416	91,34±1,09	91,32±1,09	91,21±1,09	0,8741±0,014
Vgg19-DA	91,65±1,41	0,2190	91,66±1,50	91,62±1,45	91,59±1,43	0,8792±0,020
ResNet50	91,62±1,25	0,2369	91,68±1,03	91,63±1,40	91,59±1,20	0,8786±0,018
ResNet50-DA	92,89±0,86	0,2028	92,87±0,90	92,88±1,11	92,86±0,89	0,8972±0,012
InceptionV3	89,24±1,19	0,3061	89,12±1,12	89,24±1,09	88,94±1,21	0,8419±0,017
InceptionV3-DA	92,75±0,78	0,2195	92,69±0,73	92,73±0,90	92,66±0,75	0,8949±0,011
Xception	89,04±1,09	0,2846	88,86±1,05	89,06±1,48	88,63±1,10	0,8401±0,016
Xception-DA	89,30±1,12	0,2820	89,18±1,34	89,31±1,12	88,63±1,03	0,8371±0,012
DenseNet121	92,55±1,03	0,2447	92,78±0,77	92,61±0,97	92,56±0,98	0,8922±0,014
DenseNet121-DA	94,00±0,81	0,1954	94,09±0,82	94,07±0,80	94,08±0,82	0,9133±0,011

A Tabela 19 indica o melhor desempenho obtido entre os cinco cenários para cada rede. A partir desta tabela, pode-se confirmar que o melhor resultado para ResNet50 e DenseNet121 foi encontrado quando a rotação foi aplicada como uma operação de aumento de dados. As redes InceptionV3 e AlexNet, por outro lado, tiveram o melhor impacto quando a translação foi aplicada. As redes Xception, Vgg-16 e Vgg-19 tiveram a operação de espelhamento como a que gerou seu melhor desempenho. Analisando os resultados das Tabelas 18 e 19, pode-se verificar que a rede DenseNet121 obteve os melhores resultados dentre as redes avaliadas, onde nessa última tabela ela alcançou uma acurácia de 93,69%.

Os resultados da rede AlertNet e suas variações para o cenário 5 são apresentados na Tabela 20. A rede AlertNetRWD-DA possui os melhores resultados para as taxas

Tabela 19 – Resultados obtidos com *K-fold cross validation* para os melhores resultados de aumento de dados individuais. (Os melhores valores estão em negrito).

Modelo	A(%)	P(%)	R(%)	F1-score(%)	K
ResNet50-Rotação	92,78±0,68	92,82±0,66	92,78±0,69	92,74±0,60	0,8954±0,009
InceptionV3-Translação	91,14±0,58	91,27±0,44	91,14±0,52	91,14±0,49	0,8718±0,014
Xception-Espelamento	89,41±1,01	89,35±0,91	89,41±0,96	88,98±0,96	0,8454±0,014
Vgg16-Espelamento	93,17±1,71	93,16±1,89	93,17±1,70	93,07±1,81	0,9010±0,025
Vgg19-Espelamento	92,21±0,78	92,19±0,66	92,21±0,80	92,08±0,75	0,8871±0,011
AlexNet-Translação	84,69±0,73	83,65±2,09	84,13±1,02	83,13±0,88	0,7746±0,010
DenseNet121-Rotação	93,69±0,66	93,75±0,58	93,70±0,62	93,70±0,63	0,9087±0,009

de acurácia, *recall*, *F1-score* e kappa, com 90,09% e 90,04%, 89,94% e 0,8558, respectivamente. O melhor desempenho para a precisão foi com a rede AlertNetWD-DA com valor de 90,16%.

Tabela 20 – Resultados obtidos para o cenário ALL vs. AML vs. HBS vs. outros tipos com e sem aumento de dados aplicando *k-fold cross-validation*.

Modelo	A(%)	loss	P(%)	R(%)	F1-score(%)	K
AlertNet	87,43±0,92	0,4012	87,10±1,00	87,43±0,92	86,68±0,86	0,8159±0,013
AlertNet-DA	87,97±1,31	0,3174	88,26±1,22	87,97±1,31	87,36±1,40	0,8241±0,019
AlertNetWD	85,99±0,77	0,4661	85,63±0,64	85,98±0,77	85,33±0,77	0,7948±0,011
AlertNetWD-DA	89,84±1,36	0,2748	90,16±1,31	89,83±1,36	89,84±1,35	0,8528±0,019
AlertNetR	87,54±0,65	0,4070	87,42±0,62	87,54±0,66	86,83±0,88	0,8177±0,010
AlertNetR-DA	88,28±0,75	0,3120	88,29±0,90	88,28±0,75	87,20±1,02	0,8280±0,011
AlertNetRWD	84,60±1,22	0,4922	83,87±1,29	84,60±1,22	83,47±1,18	0,7736±0,017
AlertNetRWD-DA	90,09±1,01	0,2770	90,22±0,92	90,04±0,98	89,94±1,07	0,8558±0,014
AlertNetX	86,27±1,10	0,3991	86,35±1,24	86,27±1,10	85,33±1,50	0,7986±0,016
AlertNetX-DA	86,58±1,69	0,3582	86,63±2,03	86,58±1,69	85,78±2,11	0,8034±0,025
AlertNetXWD	85,90±1,20	0,4400	85,52±1,16	85,81±1,02	85,34±1,23	0,7938±0,017
AlertNetXWD-DA	89,27±1,23	0,2964	89,34±1,37	89,27±1,22	89,02±1,47	0,8439±0,018
AlertNetRX	85,11±0,68	0,0898	84,93±1,23	85,11±0,68	84,02±1,19	0,7811±0,010
AlertNetRX-DA	85,48±1,21	0,4003	85,85±0,98	85,48±1,20	84,13±1,49	0,7862±0,018
AlertNetRXWD	83,87±1,25	0,4575	83,32±1,47	83,86±1,26	83,07±1,38	0,7634±0,018
AlertNetRXWD-DA	87,32±1,01	0,3530	87,83±1,12	87,32±1,01	86,56±1,10	0,8141±0,015

Para os cenários multiclases, os desempenhos de todas as redes analisadas foram inferiores aos cenários de classificações binárias. Um fator que contribuiu para esta diminuição nos resultados foram as classes desbalanceadas, principalmente o último cenário onde a classe “outros tipos” possui 243 imagens e a classe HBS possui 1.434 imagens. Realizando um estudo investigativo sobre o motivo das taxas terem diminuído, constatou-se que esse desbalanceamento força a rede a aprender mais sobre uma classe do que outra, visto que a classe que possui menos imagens não possui tantos exemplos para a rede treinar e aprender sobre ela. Um ponto que não alterou em relação aos cenários binário foi a melhora da performance das redes com a utilização das técnicas de aumento de dados. Assim, visto que utilizando métodos conceituados da literatura nossos resultados melhoraram, pesquisamos outros métodos para aperfeiçoar nosso estudo. Com isso, encontramos técnicas que melhoram a taxa de variância nos resultados utilizando formas de combinações de redes ou com suas predições para a classificação. A

seguir são apresentados os resultados obtidos com esses métodos investigados.

5.2.3 Multilevel e Comitê

A segunda etapa do nosso trabalho foi avaliar se a combinação das CNNs traria um benefício ao sistema de classificação. Assim, CNNs multilevel e um comitê de CNNs foram implementados. Conforme explicado anteriormente, quanto ao conceito de multilevel, as camadas de *flatten* das redes pré-treinadas são concatenadas para formar uma nova rede com várias características. Foi avaliada a arquitetura multilevel com a configuração de duas redes. Portanto, para todos os cinco cenários foram realizados testes com a combinação de duas redes. A Tabela 21 indica as combinações de redes estudadas que levaram aos melhores resultados para cada cenário em estudo

Tabela 21 – Resultados obtidos com *K-fold cross validation* para os melhores resultados com multilevel. (Os melhores resultados estão em negrito.)

Modelo-multilevel	A(%)	P(%)	R(%)	F1-score(%)	K
Leucemia - HBS					
DenseNet121-DA e Vgg16-DA	99,34±0,12	99,28±0,37	99,61±0,31	99,44±0,10	0,9864±0,002
DenseNet121-DA e InceptionV3-DA	99,31±0,06	99,36±0,13	99,52±0,24	99,44±0,06	0,9862±0,001
DenseNet121-DA e Vgg19-DA	99,39±0,09	99,28±0,23	99,75±0,16	99,52±0,08	0,9882±0,002
DenseNet121-DA e AlertNetRWD-DA	99,43±0,26	99,44±0,26	99,43±0,26	99,43±0,26	0,9882±0,005
ALL - HBS					
DenseNet121-DA e Vgg16-DA	99,70±0,24	99,85±0,30	99,65±0,24	99,75±0,19	0,9935±0,005
DenseNet121-DA and InceptionV3-DA	99,61±0,25	99,72±0,28	99,62±0,21	99,68±0,29	0,9913±0,004
ResNet50-DA and Vgg16-DA	99,69±0,25	99,73±0,31	99,63±0,23	99,74±0,20	0,9932±0,006
DenseNet121-DA e AlertNetRWD-DA	99,69±0,28	99,54±0,72	99,65±0,50	99,60±0,38	0,9935±0,006
AML - HBS					
ResNet50-DA e Vgg16-DA	99,70±0,18	99,82±0,15	99,64±0,19	99,75±0,15	0,9939±0,003
DenseNet121-DA e InceptionV3-DA	99,67±0,10	99,66±0,11	99,66±0,11	99,66±0,11	0,9931±0,002
DenseNet121-DA e Vgg16-DA	99,67±0,10	99,66±0,11	99,66±0,11	99,66±0,11	0,9931±0,002
DenseNet121-DA e AlertNetRWD-DA	99,54±0,26	98,98±0,61	99,68±0,45	99,43±0,33	0,9905±0,005
ALL - AML - HBS					
DenseNet121-DA e Vgg19-DA	97,60±0,39	97,74±0,56	97,59±0,38	97,59±0,39	0,9631±0,006
DenseNet121-DA e Vgg16-DA	97,48±0,57	97,49±0,57	97,50±0,57	97,47±0,58	0,9612±0,008
DenseNet121-DA and InceptionV3-DA	97,20±0,68	97,53±0,54	97,21±0,62	97,47±0,39	0,9601±0,007
DenseNet121-DA e AlertNetRWD-DA	96,95±0,32	97,01±0,62	96,97±0,51	97,00±0,40	0,9565±0,006
ALL - AML - HBS - outros tipos					
DenseNet121-DA e Vgg16-Espelhamento	94,73±0,61	94,75±0,60	94,72±0,62	94,71±0,62	0,9237±0,008
DenseNet121-DA e InceptionV3-DA	94,56±0,32	94,59±0,36	94,50±0,42	94,55±0,33	0,9214±0,004
DenseNet121-DA e Vgg16-DA	94,42±0,43	94,51±0,32	94,42±0,42	94,43±0,39	0,9193±0,006
DenseNet121-DA e AlertNetRWD-DA	93,90±0,41	93,99±0,64	93,97±0,58	93,96±0,50	0,9092±0,008

Comparando os resultados da Tabela 21 com os resultados individuais obtidos por cada CNN apresentados nas tabelas anteriores, percebe-se que o Multilevel melhorou os resultados apenas nos problemas de classificação multiclasse. Investigando o motivo dessa melhoria encontrou-se no trabalho de [Srinivasan, Dubey e Ganeshbabu \(2016\)](#) que, sem o uso de técnicas de combinação em conjunto de dados não balanceados, geralmente são obtidos desempenhos de classificação inferiores. O que foi o caso nos problemas multiclasse aqui estudados, pois existem classes desbalanceadas, principalmente, para o último cenário. Assim, segundo os autores, o uso de uma abordagem de classificação multilevel pode resolver os problemas de desbalanceamento de dados e reduzir a variação dos erros de estimativa.

Ainda, realizando experimentos com técnica de combinação, também é aplicado e avaliado o método comitê de classificadores. Tal como acontece com a técnica multilevel, foi feita uma análise dos resultados anteriores para formar um comitê e melhorar o desempenho da classificação. Assim, foram selecionadas as redes que obtiveram os melhores desempenhos e realizadas as combinações de CNNs pré-treinadas como entrada para o modelo de comitê composto por três, quatro, cinco e seis redes. As redes AlertNet e suas variações não estão entre as redes que apresentaram os melhores resultados nos cenários de classificação, por isso não estão compondo os comitês testados. A Tabela 22 apresenta os resultados encontrados entre os testes realizados para cada cenário em estudo. Para os cinco cenários, o desempenho melhorou, tanto em termos de valores métricos quanto de desvio padrão.

Tabela 22 – Resultados obtidos com *K-fold cross validation* para os melhores resultados com comitê.

Modelo-comitê	A(%)	P(%)	R(%)	F1-score(%)	K
Leucemia - HBS					
DenseNet121-DA, ResNet50-DA e ResNet50	99,54±0,23	99,54±0,23	99,54±0,23	99,54±0,23	0,9905±0,004
DenseNet121-DA, ResNet50-DA, DenseNet121 e ResNet50	99,54±0,23	99,54±0,23	99,54±0,23	99,54±0,23	0,9905±0,004
DenseNet121-DA, ResNet50-DA, DenseNet121, ResNet50 e Vgg16-DA	99,43±0,14	99,43±0,14	99,43±0,14	99,43±0,14	0,9882±0,002
DenseNet121-DA, ResNet50-DA, DenseNet121, ResNet50, Vgg16-DA e Vgg19-DA	99,45±0,11	99,45±0,12	99,46±0,11	99,45±0,11	0,9887±0,002
ALL - HBS					
DenseNet121-DA, ResNet50-DA e DenseNet121	99,82±0,28	99,82±0,27	99,82±0,28	99,82±0,28	0,9963±0,006
DenseNet121-DA, ResNet50-DA, DenseNet121 e ResNet50	99,82±0,28	99,82±0,28	99,82±0,28	99,82±0,28	0,9963±0,006
DenseNet121-DA, ResNet50-DA, DenseNet121, ResNet50 e Vgg19-DA	99,91±0,12	99,91±0,12	99,90±0,13	99,91±0,12	0,9981±0,002
DenseNet121-DA, ResNet50-DA, DenseNet121, ResNet50, Vgg19-DA e Vgg16-DA	99,95±0,09	99,95±0,09	99,95±0,10	99,95±0,09	0,9990±0,002
AML - HBS					
DenseNet121-DA, ResNet50-DA e Vgg16-DA	99,79±0,14	99,79±0,14	99,79±0,14	99,79±0,14	0,9957±0,003
DenseNet121-DA, ResNet50-DA, InceptionV3-DA e Vgg16-DA	99,87±0,11	99,87±0,11	99,87±0,11	99,87±0,11	0,9974±0,002
DenseNet121-DA, ResNet50-DA, ResNet50, InceptionV3-DA e Vgg16-DA	99,74±0,09	99,74±0,09	99,74±0,09	99,74±0,09	0,9948±0,002
DenseNet121-DA, ResNet50-DA, ResNet50, InceptionV3-DA, Vgg16-DA e Vgg19-DA	99,87±0,11	99,87±0,11	99,87±0,11	99,87±0,11	0,9974±0,002
ALL - AML - HBS					
DenseNet121-DA, Vgg16-DA e Vgg19-DA	97,35±0,57	97,44±0,53	97,36±0,57	97,35±0,56	0,9593±0,008
DenseNet121-DA, ResNet50-DA, Vgg16-DA and Vgg19-DA	97,35±0,69	97,41±0,67	97,35±0,69	97,35±0,68	0,9593±0,010
DenseNet121-DA, ResNet50-DA, Vgg16-DA, Vgg19-DA e InceptionV3-DA	97,44±0,84	97,54±0,76	97,45±0,83	97,44±0,83	0,9607±0,012
DenseNet121-DA, ResNet50-DA, Vgg16-DA, Vgg19-DA, InceptionV3-DA e Vgg16	97,53±0,62	97,61±0,62	97,54±0,62	97,53±0,62	0,9621±0,009
ALL - AML - HBS - outros tipos					
DenseNet121-DA, DenseNet121-Translação e DenseNet121-Rotação	94,42±0,30	94,40±0,34	94,41±0,32	94,42±0,32	0,9194±0,004
DenseNet121-DA, DenseNet121-Translação, DenseNet121-Espelhamento e DenseNet121-Rotação	93,82±0,47	94,02±0,47	93,89±0,46	93,86±0,49	0,9104±0,006
DenseNet121-DA, DenseNet121-Rotação, DenseNet121-Translação, DenseNet121-Espelhamento e DenseNet121-Zoom	94,59±0,77	94,65±0,72	94,60±0,78	94,58±0,79	0,9218±0,011
DenseNet121-DA, DenseNet121-Rotação, DenseNet121-Translação, DenseNet121-Espelhamento, DenseNet121-Zoom e Vgg16-Espelhamento	94,45±0,83	94,54±0,77	94,48±0,81	94,45±0,83	0,9196±0,012

Observa-se que selecionar CNNs diversificadas que apresentam altas taxas de precisão em várias regiões no espaço de características é essencial para a construção de

configurações multilevel e comitês. Por esse motivo, foram avaliadas as combinações de modelos e determinada a melhor para construir a configuração do comitê proposto. Combinar métodos CNNs não apenas aumenta o desempenho, mas também reduz o risco de *overfitting* (SOLLICH; KROGH, 1995). Os resultados experimentais obtidos pelas redes desenvolvidas neste estudo são estatisticamente significativos, principalmente quando comparadas com as redes pré-treinadas. Com isso, pressupõe que a abordagem proposta pode ajudar a desenvolver soluções clinicamente valiosas para detectar e diferenciar imagens de leucemia em lâminas de sangue.

Para uma melhor visualização do desempenho dos métodos aplicados, a Figura 18 apresenta uma representação gráfica dos melhores resultados encontrados para os experimentos com CNNs individuais, com CNNs multilevel e configurações de comitê quanto aos cinco cenários avaliados em estudo. Nesta figura, podemos visualizar uma melhora dos resultados alcançados com a técnica de comitê ou multilevel sobre os melhores resultados individuais das redes para quase todos os cenários. Pode-se perceber também que para os problemas multiclasse (Figura 18 (d) e 18 (e)), as técnicas de combinação obtiveram uma melhora significativa quanto ao melhor desempenho individual das CNNs. Outro ponto a ser destacado é o desempenho individual das redes DenseNet121-DA e ResNet50-DA, que obtiveram os melhores resultados nos cinco cenários em relação às demais redes avaliadas.

5.2.4 Discussões

Nesta seção é apresentada uma análise comparativa dos resultados obtidos pelo método proposto em relação aos trabalhos relacionados (Capítulo 2).

A Tabela 23 apresenta uma comparação com artigos do estado da arte quanto ao problema de classificação, número de conjuntos de dados, número de imagens e acurácia. Os resultados sugerem que mesmo usando CNNs de uso geral, a escolha das técnicas corretas de aumento de dados juntamente com uma combinação adequada de CNNs, demonstra que resultados que são competitivos com os métodos de última geração podem ser alcançados. Para fazer uma comparação mais confiável, a tabela foi dividida de acordo com o tipo de classificação utilizada nos estudos. Assim, para a avaliação entre Leucemia e HBS, encontram-se excelentes resultados, como o trabalho de Loey, Naman e Zayed (2020) que atingiu 100% de acurácia. Podemos notar que as configurações de comitê e multilevel utilizadas neste estudo também obtiveram ótimos resultados. No entanto, embora seja aplicada a rede AlexNet nos nossos testes, como sugerido no trabalho de Loey, Naman e Zayed (2020), essa rede não obteve um bom desempenho, o que pode ser devido à profundidade da rede AlexNet ser muito menor do que as outras redes estudadas. Portanto, a rede apresentou dificuldade para aprender as características dos conjuntos de dados de imagens usados.

Figura 18 – Comparação dos melhores resultados obtidos pelas configurações de CNNs individuais, multilevel e comitê no cenários (a) Leucemia vs. HBS, (b) ALL vs. HBS, (c) AML vs. HBS, (d) ALL vs. AML vs. HBS and (e) ALL vs. AML vs. HBS vs. outros tipos.

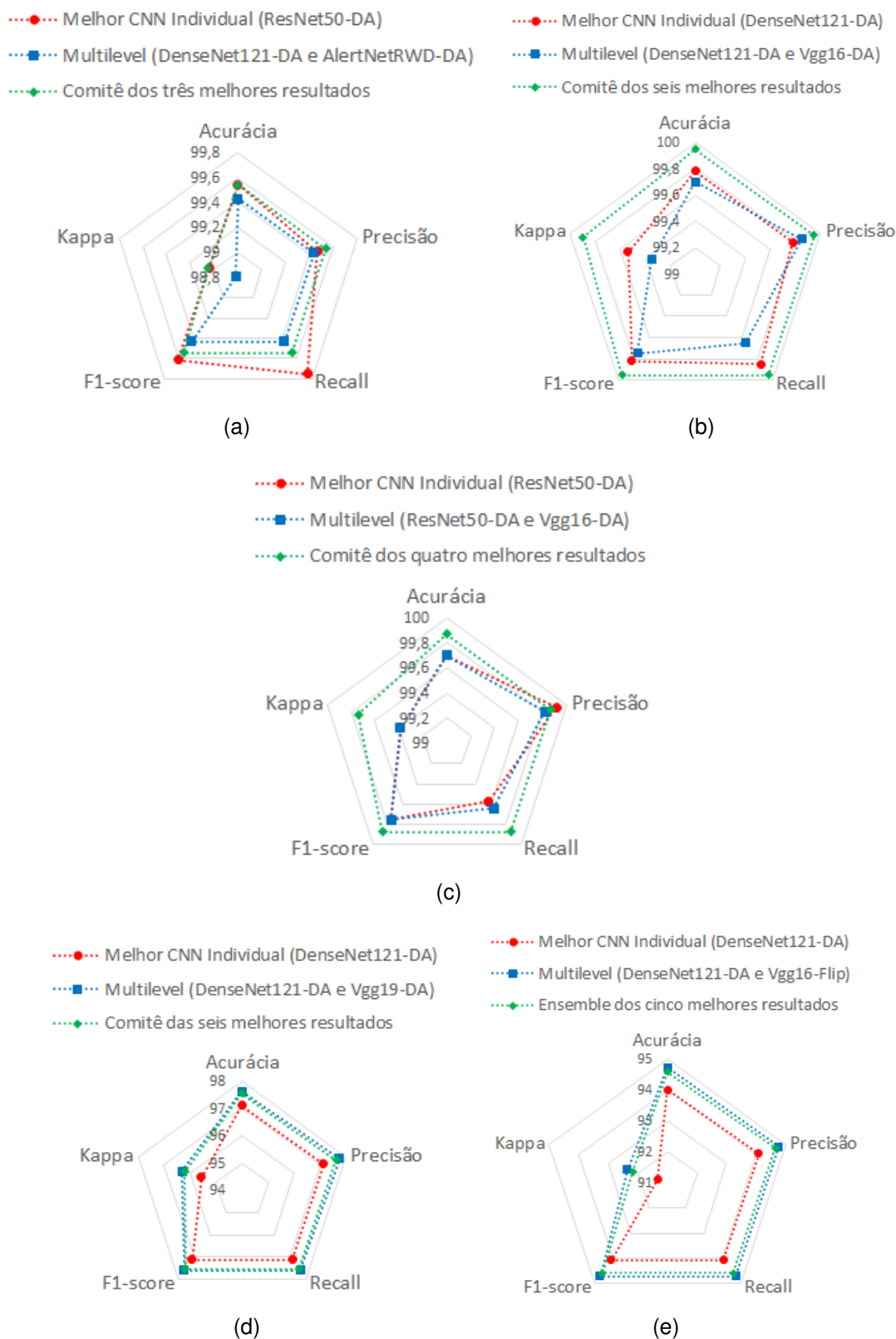


Tabela 23 – Comparação entre os resultados obtidos pelas configurações multilevel e comitê propostas com os obtidos pelos métodos relacionados.

Método	Classificação	Número de bases de imagens	Número de imagens	Acurácia(%)
Thanh et al. (2018)	Leucemia - HBS	1	108	96,60
Vogado et al. (2018)	Leucemia - HBS	8	1,268	99,76
Loey, Naman e Zayed (2020)	Leucemia - HBS	2	564	100
Vogado et al. (2021)	Leucemia - HBS	18	3.536	98,61
Multilevel	Leucemia - HBS	18	3.536	99,43
Comitê	Leucemia - HBS	18	3.536	99,54
Singhal e Singh (2016)	ALL - HBS	1	260	93,80
Shafique e Tehsin (2018)	ALL - HBS	2	368	99,50
Ahmed et al. (2019)	ALL - HBS	2	354	88,25
Pansombut et al. (2019)	ALL - HBS	2	363	81,74
Gehlot, Gupta e Gupta (2020)	ALL - HBS	1	15.114	93,40
Das e Meher (2021)	ALL - HBS	2	368	97,92 e 96,00
Khandekar et al. (2021)	ALL - HBS	2	1.108	mAP:96,06;98,70
Ullah et al. (2021)	ALL - HBS	1	15.114	91,10
Karar, Alotaibi e Alotaibi (2022)	ALL - HBS	1	368	98,65
Rodrigues et al. (2022)	ALL - HBS	1	260	98,46
Abhishek et al. (2022)	ALL - HBS	2	608	97,00
Rastogi, Khanna e Singh (2022)	ALL - HBS	1	260	96,15
Multilevel	ALL - HBS	14	2.315	99,70
Comitê	ALL - HBS	14	2.315	99,95
Madhukar, Agaian e Chronopoulos (2012)	AML - HBS	1	50	93,50
Goutam e Sailaja (2015)	AML - HBS	1	90	98,00
Dasariraju, Huo e McCalla (2020)	AML - HBS	1	1.274	92,99
Multilevel	AML - HBS	16	2.412	99,70
Comitê	AML - HBS	16	2.412	99,87
Rawat et al. (2017)	ALL - AML - HBS	1	240	99,50
Laosai e Chamnongthai (2018)	ALL - AML - HBS	2	500	99,85
Tran et al. (2018)	ALL - AML - HBS	2	141	97,30
Karar, Alotaibi e Alotaibi (2022)	ALL - AML - HBS	2	445	95,50
Abhishek et al. (2022)	ALL - AML - HBS	2	608	98,00
Ghaderzadeh et al. (2022)	ALL - AML - HBS	1	12.528	98,50
Multilevel	ALL - AML - HBS	18	3.293	97,60
Comitê	ALL - AML - HBS	18	3.293	97,53
Multilevel	ALL - AML - HBS - outros tipos	18	3.536	94,73
Comitê	ALL - AML - HBS - Outros tipos	18	3.536	94,59

As configurações de multilevel e de comitê obtiveram melhor desempenho na classificação de imagens tanto para ALL e HBS quanto para AML e HBS. A configuração do comitê obteve 99,95% de acurácia para a primeira classificação e 99,87% para a segunda. Percebe-se que os demais trabalhos que utilizaram esses classificadores não estudaram uma diversidade de imagens muito grande, pois a maioria deles usou apenas um conjunto de dados.

Quanto à classificação em três classes (ALL-AML-HBS), o estudo de Laosai e Chamnongthai (2018) obteve o melhor resultado com 99,85% de acurácia em 500 imagens de dois conjuntos de dados privados. Para os testes realizados neste trabalho foram utilizadas 3.293 imagens publicamente disponíveis, o que é quase sete vezes maior do que o conjunto de dados usado por Loey, Naman e Zayed (2020), as configurações de multilevel e comitê atingiram 97,60% e 97,53% de acurácia, respectivamente.

Percebe-se que a maioria dos trabalhos do estado da arte apresentou valores de acertos superiores aos da abordagem proposta. No entanto, o número de imagens utilizadas por estes estudos, quando comparadas ao trabalho executado, são bem menores. Outro detalhe a ser destacado é que a maioria dos artigos relacionados utilizou técnicas de aumento de dados para reduzir o *overfitting*. Também é importante ressaltar que foi utilizado um conjunto variável de imagens de dezoito conjuntos de dados disponíveis publicamente com diferentes características de imagem como iluminação, contraste e brilho, o que levou a uma maior diversidade nos dados de treinamento, permitindo a obtenção de um classificador mais robusto para diferentes imagens de

entrada.

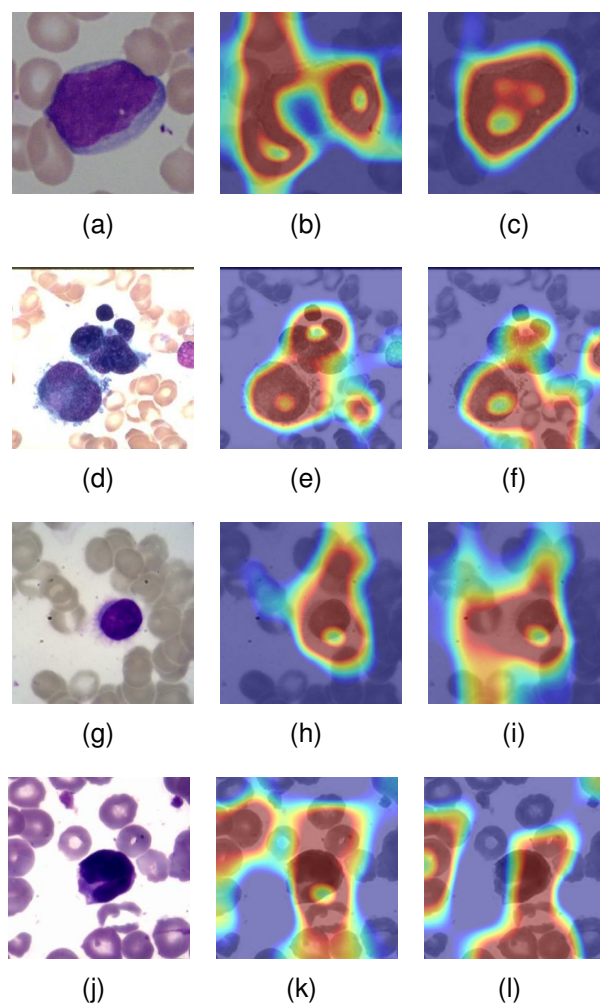
Até o presente momento, a literatura não dispõe de trabalhos que utilizaram a divisão ALL vs. AML vs. HBS vs. outros tipos mencionados anteriormente, sendo um dos diferenciais deste estudo a abordagem dos quatro tipos de leucemia. De acordo com a Tabela 23, foi obtido 94,73% de acurácia pela configuração multilevel e 94,59% pela configuração de comitê. Não foi aplicada a divisão na leucemia crônica (CML e CLL), porque o número de imagens para a classe “outros tipos” já era menor do que as outras três classes estudadas: ALL, AML e HBS. Então, se ainda fosse fazer essa subdivisão, as classes ficariam ainda mais desbalanceadas, mesmo após a aplicação de técnicas de aumento de dados.

A Figura 19 apresenta exemplos de mapas de ativação obtidos pelas redes Vgg16 e DenseNet121 para o cenário de quatro classes. A primeira coluna apresenta as imagens originais, a segunda coluna corresponde aos resultados obtidos na rede Vgg16 e a terceira coluna apresenta os resultados obtidos com o DenseNet121. Para estas redes, os tons de azul indicam uma baixa ativação, sugerindo que as regiões correspondentes são de baixa importância para a classificação final; inversamente, os tons vermelhos estão associados às áreas mais críticas para a classificação final. Essas regiões do esfregaço de sangue periférico são de grande importância para o clínico que examina a aparência das células, pois as alterações no número e na aparência das células frequentemente ajudam no diagnóstico de leucemia.

Uma alternativa não avaliada neste trabalho para obter melhores resultados seria aprimorar a imagem antes de iniciar o processo de classificação. Por exemplo, uma mudança de contraste adequada pode melhorar a qualidade perceptiva da maioria das imagens (GU et al., 2015). Na verdade, de acordo com Gu et al. (2016), um bom aprimoramento pode melhorar visivelmente a qualidade das imagens de entrada, ainda melhor do que as imagens capturadas originalmente, que geralmente são consideradas da melhor qualidade. No entanto, a escolha do método de aprimoramento de imagem deve ser prudente, uma vez que a maioria das tecnologias relevantes muitas vezes sofre da desvantagem do aprimoramento excessivo, introduzindo ruído/artefatos e alterando as regiões de atenção visual (GU et al., 2018).

Outro ponto essencial a ser avaliado é a utilização de CNNs pré-treinadas com ImageNet. De acordo com Kornblith, Shlens e Le (2019), as arquiteturas de melhor desempenho no ImageNet podem fornecer melhor extração e ajuste fino de características. No entanto, os autores observaram esse fato apenas em conjuntos de dados fotográficos. De acordo com Sipes e Li (2018), as imagens de leucemia são consideradas de baixa granularidade. As tarefas de classificação em conjuntos de dados com imagens refinadas podem ser particularmente desafiadoras, e os efeitos do pré-treinamento com ImageNet foram considerados pequenos. Portanto, é essencial

Figura 19 – Exemplos de mapas de ativação para imagens de lâminas de sangue - 1º coluna: imagens originais; 2º coluna: Rede Vgg16 e; 3º coluna: Rede DenseNet121. (a-c) imagens ALL, (d-f) imagens AML, (g-i) imagens com Leucemia Crônica e (j-l) imagens HBS.



avaliar o uso de técnicas de refinamento específicas para imagens de granulação fina (DU et al., 2021).

Para os testes realizados, são utilizadas as arquiteturas CNNs originais. No entanto, mudanças nessas arquiteturas podem trazer melhorias de desempenho. Por exemplo, a inserção de mecanismos de atenção como o proposto por Xie et al. (2021). A ideia central dos mecanismos de atenção é cada vez que o modelo prevê uma saída, ele usa apenas partes da entrada onde as informações mais relevantes estão concentradas.

Este trabalho avaliou as combinações multilevel e de comitê, mas existem outras configurações para métodos baseados em CNN. Por exemplo, estruturas hierárquicas podem ser usadas para problemas com várias classes. Esses modelos exploram a estrutura hierárquica das categorias de objetos para decompor as tarefas de classificação em várias etapas. Eles mostraram melhor desempenho do que os modelos planos na classificação de imagens em vários domínios (KOWSARI et al., 2020).

5.3 Considerações Finais

Os métodos apresentados neste trabalho apresentam excelentes resultados de acordo com o índice *Kappa* para a detecção de leucemia em cinco cenários de classificação. Quando comparados os resultados das técnicas de multilevel e comitê com os resultados individuais obtidos pelas redes, pode-se perceber uma diminuição nas taxas de desvio padrão e aumento das taxas de acurácia, *recall*, precisão, *f1-score* e *kappa*, principalmente nas classificações multiclasse.

6 Conclusões e Trabalhos Futuros

6.1 Conclusões

Nesta tese, foram avaliadas técnicas que podem ser integradas aos sistemas de diagnóstico auxiliados por computador (CADs) para detectar os tipos de leucemia (ALL, AML e outros tipos), além de lâminas de sangue saudáveis. Diversos experimentos foram realizados, onde primeiro foram desenvolvidas CNNs (*Acute Leukemias Recognition Network* - AlertNet e suas variações) para detectar os tipos de leucemia aguda e imagens saudáveis em 3.293 imagens de lâminas de sangue. Logo em sequência, foram realizados testes em cinco cenários de classificação para analisar a eficácia obtida através das técnicas de aumento de dados. Em seguida, é realizada uma comparação entre as técnicas de aumento de dados na classificação das imagens em ALL vs. AML vs. HBS vs. outros tipos. O número total de imagens utilizadas foi de 3.536, obtidas a partir de diferentes conjuntos de dados públicos.

Os métodos propostos para detecção de leucemia foram aplicados em dezoito bases de imagens públicas oriundas de diferentes regiões do mundo, o que ocasiona uma boa diversidade de imagens. Assim, essa variedade de imagens faz com que a rede aprenda várias características, como por exemplo, pode existir mais de uma célula na imagem que tem leucemia, com isso ela aprende a não focar apenas em uma determinada região da imagem, pois não sabe a quantidade de células existentes e nem em qual parte da imagem as células estarão. Com esse pensamento utilizamos cinco técnicas de aumento de dados para ajudar a aumentar a variedade das imagens e diminuir o *overfitting* das redes.

O aumento de dados está se tornando uma importante área de pesquisa, especialmente para imagens médicas. Existem artigos sobre métodos de aumento de dados para tipos específicos de imagens. No entanto, não há trabalho abrangente suficiente sobre o aumento das imagens e classificações de leucemia. Portanto, este estudo irá beneficiar muitos pesquisadores que trabalham no mesmo assunto.

Com base nas redes pré-treinadas da literatura, propusemos a *Acute Leukemias Recognition Network* (AlertNet) e suas variações para classificar os tipos de leucemia. Essas redes possuem uma menor quantidade de parâmetros quando comparados com redes como a DenseNet121 e Xception. Inicialmente essa rede foi aplicada apenas para a diferenciação entre os dois tipos de leucemia aguda (ALL-AML) e imagens saudáveis, onde foram obtidos resultados promissores com uma acurácia e precisão de 96,17% e 96,38%, respectivamente.

A partir desses resultados iniciais podemos inferir que, ao se utilizar o *dropout* é necessário aumentar o tamanho do modelo pois, normalmente, o erro do conjunto de dados de validação é muito menor ao utilizá-lo, mas modelos maiores custam mais iterações de treinamento. Quando o conjunto de dados de treinamento é pequeno, o uso do *dropout* se torna menos eficaz. Por isso que os resultados encontrados neste estudo sem essa técnica foram melhores do que quando a utilizamos.

Logo após essa parte inicial, realizamos novos experimentos aplicando as redes propostas para outros tipos de classificação, o que totalizou cinco cenários ao todo, três de classificação binária (Leucemia vs. HBS / ALL vs. HBS / AML vs. HBS) e dois cenários de classificação multiclasse (ALL vs. AML vs. HBS / ALL vs. AML vs. HBS vs. outros tipos). Além das redes propostas, também aplicamos sete rede pré-treinadas, sendo elas as redes AlexNet, Vgg16, Vgg19, ResNet50, InceptionV3, Xception e DenseNet121.

Foram feitos experimentos com estas redes utilizando ou não o aumento de dados para os cinco cenários. A partir dos desempenhos individuais das redes, avaliamos e utilizamos os melhores resultados para compor o multilevel e comitê. Esses conceitos foram aplicados nos experimentos conduzidos, com o objetivo de aumentar a precisão e generalização da abordagem proposta e diminuir a taxa de desvio padrão e, consequentemente, a variância. Para os cinco cenários de classificação dos tipos de leucemia foram obtidos excelentes resultados com a aplicação dessas técnicas.

O comitê obteve melhor desempenho nas classificações binárias com 99,54% para a classificação entre leucemia vs. HBS, utilizando o voto majoritário com as previsões das redes. Para o cenário ALL vs. HBS foi alcançado uma taxa de acurácia de 99,95% e o cenário AML vs. HBS o melhor resultado foi de 99,87% de acurácia. Para os dois cenários multiclasse os melhores desempenhos do nosso estudo foi com a técnica de multilevel, onde ela obteve 97,60% e 94,73% de acurácia para os cenários ALL vs. AML vs. HBS e ALL vs. AML vs. HBS vs. outros tipos, respectivamente.

É imprescindível destacar o uso da configuração de comitê, pois reduz estas taxas no modelo de classificação, ao combinar de forma otimizada as previsões de vários modelos. O desempenho do modelo de comitê simula condições do mundo real com desvio padrão e ajuste excessivo reduzido, levando a uma generalização aprimorada. Assim, pode-se concluir que os resultados obtidos podem auxiliar no desenvolvimento de soluções clinicamente valiosas para detectar e diferenciar os tipos de leucemia em lâminas de sangue.

Assim, os resultados sugerem que mesmo usando CNNs de uso geral, a escolha das técnicas corretas de aumento de dados juntamente com uma combinação adequada de CNNs, demonstra que os resultados encontrados podem ser alcançados quando equiparados com os métodos de última geração.

6.2 Trabalhos Futuros

Apesar dos bons resultados obtidos para a classificação dos tipos de leucemia em imagens de lâminas de sangue, melhorias ainda podem ser feitas no método proposto visando incrementar a sua eficiência. A seguir são destacadas algumas sugestões:

1. Considerando a diversidade de aquisições das bases de imagens, padronizar as entradas da rede usando uma rede autoencoder como um pré-processamento poderia aumentar a eficiência dos modelos no treinamento da rede;
2. Como houve uma melhora nos resultados utilizando multilevel e ensemble, investigar e avaliar novas técnicas de combinação visando a melhora no desempenho;
3. Como existe algumas imagens de leucemia crônica neste estudo, realizar a diferenciação entre imagens de células com leucemia linfóide crônica e leucemia mieloide crônica;
4. Aplicar técnicas para balancear as classes a fim de melhorar os resultados.

Referências

ABADI, M.; AGARWAL, A.; BARHAM, P.; BREVDO, E.; CHEN, Z.; CITRO, C.; CORRADO, G. S.; DAVIS, A.; DEAN, J.; DEVIN, M. et al. Tensorflow: Large-scale machine learning on heterogeneous distributed systems. **arXiv preprint arXiv:1603.04467**, 2016. Nenhuma citação no texto.

ABBAS, N.; MOHAMAD, D. Automatic color nuclei segmentation of leukocytes for acute leukemia. **Research Journal of Applied Sciences, Engineering and Technology**, Maxwell Science Publishing, v. 7, n. 14, p. 2987–2993, 2014. Nenhuma citação no texto.

ABHISHEK, A.; JHA, R. K.; SINHA, R.; JHA, K. Automated classification of acute leukemia on a heterogeneous dataset using machine learning and deep learning techniques. **Biomedical Signal Processing and Control**, Elsevier, v. 72, p. 103341, 2022. Nenhuma citação no texto.

AGAIAN, S.; MADHUKAR, M.; CHRONOPOULOS, A. T. A new acute leukaemia-automated classification system. **Computer Methods in Biomechanics and Biomedical Engineering: Imaging & Visualization**, Taylor & Francis, v. 6, n. 3, p. 303–314, 2018. Nenhuma citação no texto.

AHMED, M. J.; NAYAK, P. A survey on leukemia detection using image processing techniques. In: EDP SCIENCES. **E3S Web of Conferences**. [S.I.], 2021. v. 309. Nenhuma citação no texto.

AHMED, N.; YIGIT, A.; ISIK, Z.; ALPKOCAK, A. Identification of leukemia subtypes from microscopic images using convolutional neural network. **Diagnostics**, Multidisciplinary Digital Publishing Institute, v. 9, n. 3, p. 104, 2019. Nenhuma citação no texto.

ANILKUMAR, K.; MANOJ, V.; SAGI, T. A survey on image segmentation of blood and bone marrow smear images with emphasis to automated detection of leukemia. **Biocybernetics and Biomedical Engineering**, Elsevier, v. 40, n. 4, p. 1406–1420, 2020. Nenhuma citação no texto.

ANILKUMAR, K.; MANOJ, V.; SAGI, T. Automated detection of leukemia by pretrained deep neural networks and transfer learning: a comparison. **Medical Engineering & Physics**, Elsevier, v. 98, p. 8–19, 2021. Nenhuma citação no texto.

ARSLAN, S.; OZYUREK, E.; GUNDUZ-DEMIR, C. A color and shape based algorithm for segmentation of white blood cells in peripheral blood and bone marrow images. **Cytometry Part A**, Wiley Online Library, v. 85, n. 6, p. 480–490, 2014. Nenhuma citação no texto.

BELKIN, M.; HSU, D.; MA, S.; MANDAL, S. Reconciling modern machine-learning practice and the classical bias–variance trade-off. **Proceedings of the National Academy of Sciences**, National Acad Sciences, v. 116, n. 32, p. 15849–15854, 2019. Nenhuma citação no texto.

BIBI, N.; SIKANDAR, M.; DIN, I. U.; ALMOGREN, A.; ALI, S. Iomt-based automated detection and classification of leukemia using deep learning. **Journal of Healthcare Engineering**, Hindawi, v. 2020, 2020. Nenhuma citação no texto.

BÖHM, J. Pathologie-websites im world wide web. **Der Pathologe**, v. 29, n. 3, p. 231–242, 2008. Nenhuma citação no texto.

CHATFIELD, K.; SIMONYAN, K.; VEDALDI, A.; ZISSERMAN, A. Return of the devil in the details: Delving deep into convolutional nets. **arXiv preprint arXiv:1405.3531**, 2014. Nenhuma citação no texto.

CHEN, X. **Image enhancement effect on the performance of convolutional neural networks**. 2019. Nenhuma citação no texto.

CHOLLET, F. Xception: Deep learning with depthwise separable convolutions. In: **IEEE Conference on Computer Vision and Pattern Recognition**. [S.l.: s.n.], 2017. p. 1800–1807. Nenhuma citação no texto.

DAHL, G. E.; SAINATH, T. N.; HINTON, G. E. Improving deep neural networks for lvcsr using rectified linear units and dropout. In: IEEE. **IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)**. [S.l.], 2013. p. 8609–8613. Nenhuma citação no texto.

DAS, P. K.; MEHER, S. An efficient deep convolutional neural network based detection and classification of acute lymphoblastic leukemia. **Expert Systems with Applications**, Elsevier, v. 183, p. 115311, 2021. Nenhuma citação no texto.

DASARATHY, B. V.; SHEELA, B. V. A composite classifier system design: Concepts and methodology. **Proceedings of the IEEE**, IEEE, v. 67, n. 5, p. 708–713, 1979. Nenhuma citação no texto.

DASARIRAJU, S.; HUO, M.; MCCALLA, S. Detection and classification of immature leukocytes for diagnosis of acute myeloid leukemia using random forest algorithm. **Bioengineering**, Multidisciplinary Digital Publishing Institute, v. 7, n. 4, p. 120, 2020. Nenhuma citação no texto.

DENG, J.; DONG, W.; SOCHER, R.; LI, L.-J.; LI, K.; LI, F.-F. Imagenet: a large-scale hierarchical image database. In: . [S.l.: s.n.], 2009. p. 248–255. Nenhuma citação no texto.

DIETTERICH, T. G. Ensemble methods in machine learning. In: SPRINGER. **International workshop on multiple classifier systems**. [S.l.], 2000. p. 1–15. Nenhuma citação no texto.

DU, R.; XIE, J.; MA, Z.; CHANG, D.; SONG, Y.-Z.; GUO, J. Progressive learning of category-consistent multi-granularity features for fine-grained visual classification. **IEEE Transactions on Pattern Analysis and Machine Intelligence**, p. 1–1, 2021. Nenhuma citação no texto.

DWIVEDI, A. K. Artificial neural network model for effective cancer classification using microarray gene expression data. **Neural Computing and Applications**, Springer, v. 29, n. 12, p. 1545–1554, 2018. Nenhuma citação no texto.

FUHAD, K. F.; TUBA, J. F.; SARKER, M. R. A.; MOMEN, S.; MOHAMMED, N.; RAHMAN, T. Deep learning based automatic malaria parasite detection from blood smear and its smartphone based application. **Diagnostics**, MDPI, v. 10, n. 5, p. 329, 2020. Nenhuma citação no texto.

FUKUSHIMA, K. Neocognitron: A hierarchical neural network capable of visual pattern recognition. **Neural networks**, Elsevier, v. 1, n. 2, p. 119–130, 1988. Nenhuma citação no texto.

GEHLOT, S.; GUPTA, A.; GUPTA, R. Sdct-auxnet θ : Dct augmented stain deconvolutional cnn with auxiliary classifier for cancer diagnosis. **Medical Image Analysis**, Elsevier, v. 61, p. 101661, 2020. Nenhuma citação no texto.

GÉRON, A. **Hands-on machine learning with Scikit-Learn, Keras, and TensorFlow**. [S.l.]: "O'Reilly Media, Inc.", 2022. Nenhuma citação no texto.

GHADERZADEH, M.; HOSSEINI, A.; ASADI, F.; ABOLGHASEMI, H.; BASHASH, D.; ROSHANPOOR, A. Automated detection model in classification of b-lymphoblast cells from normal b-lymphoid precursors in blood smear microscopic images based on the majority voting technique. **Scientific Programming**, Hindawi, v. 2022, 2022. Nenhuma citação no texto.

GOCERI, E. Image augmentation for deep learning based lesion classification from skin images. In: **Proceedings of the IEEE 4th International Conference on Image Processing, Applications and Systems**. [S.l.: s.n.], 2020. p. 144–148. Nenhuma citação no texto.

GOODFELLOW, I.; BENGIO, Y.; COURVILLE, A. Deep learning (adaptive computation and machine learning series). **Cambridge Massachusetts**, p. 321–359, 2017. Nenhuma citação no texto.

GOUTAM, D.; SAILAJA, S. Classification of acute myelogenous leukemia in blood microscopic images using supervised classifier. In: IEEE. **2015 IEEE International Conference on Engineering and Technology (ICETECH)**. [S.l.], 2015. p. 1–5. Nenhuma citação no texto.

GU, K.; TAO, D.; QIAO, J.-F.; LIN, W. Learning a no-reference quality assessment model of enhanced images with big data. **IEEE Transactions on Neural Networks and Learning Systems**, v. 29, n. 4, p. 1301–1313, 2018. Nenhuma citação no texto.

GU, K.; ZHAI, G.; LIN, W.; LIU, M. The analysis of image contrast: From quality assessment to automatic enhancement. **IEEE Transactions on Cybernetics**, v. 46, n. 1, p. 284–297, 2016. Nenhuma citação no texto.

GU, K.; ZHAI, G.; YANG, X.; ZHANG, W.; CHEN, C. W. Automatic contrast enhancement technology with saliency preservation. **IEEE Transactions on Circuits and Systems for Video Technology**, v. 25, n. 9, p. 1480–1494, 2015. Nenhuma citação no texto.

GUPTA, A.; GUPTA, R.; GEHLOT, S.; MOURYA, S. Classification of normal vs malignant cells in b-all white blood cancer microscopic images. In: **IEEE International Symposium on Biomedical Imaging (ISBI)-2019 Challenges Internet**. [S.l.: s.n.], 2019. Nenhuma citação no texto.

HE, K.; ZHANG, X.; REN, S.; SUN, J. Deep residual learning for image recognition. In: **IEEE Conference on Computer Vision and Pattern Recognition**. [S.l.]: IEEE, 2016. p. 770–778. Nenhuma citação no texto.

HEMELINGS, R.; ELEN, B.; BARBOSA-BREDA, J.; BLASCHKO, M. B.; BOEVER, P. D.; STALMANS, I. Deep learning on fundus images detects glaucoma beyond the optic disc. **Scientific Reports**, Nature Publishing Group, v. 11, n. 1, p. 1–12, 2021. Nenhuma citação no texto.

HINTON, G. E.; SRIVASTAVA, N.; KRIZHEVSKY, A.; SUTSKEVER, I.; SALAKHUTDINOV, R. R. Improving neural networks by preventing co-adaptation of feature detectors. **arXiv preprint arXiv:1207.0580**, 2012. Nenhuma citação no texto.

HUANG, G.; LIU, Z.; MAATEN, L. V. D.; WEINBERGER, K. Q. Densely connected convolutional networks. In: **Proceedings of the IEEE conference on computer vision and pattern recognition**. [S.l.: s.n.], 2017. p. 4700–4708. Nenhuma citação no texto.

IOFFE, S.; SZEGEDY, C. Batch normalization: Accelerating deep network training by reducing internal covariate shift. 02 2015. Nenhuma citação no texto.

IZADYYAZDANABADI, M.; BELYKH, E.; MOONEY, M.; MARTIROSYAN, N.; ESCHBACHER, J.; NAKAJI, P.; PREUL, M. C.; YANG, Y. Convolutional neural networks: Ensemble modeling, fine-tuning and unsupervised semantic localization for neurosurgical cle images. **Journal of Visual Communication and Image Representation**, Elsevier, v. 54, p. 10–20, 2018. Nenhuma citação no texto.

JANOCHA, K.; CZARNECKI, W. M. On loss functions for deep neural networks in classification. **arXiv:1702.05659**, 2017. Nenhuma citação no texto.

JIA, Y.; SHELHAMER, E.; DONAHUE, J.; KARAYEV, S.; LONG, J.; GIRSHICK, R.; GUADARRAMA, S.; DARRELL, T. Caffe: Convolutional architecture for fast feature embedding. In: ACM. **Proceedings of the 22nd ACM international conference on Multimedia**. [S.l.], 2014. p. 675–678. Nenhuma citação no texto.

JIANG, K.; LIAO, Q.-M.; DAI, S.-Y. A novel white blood cell segmentation scheme using scale-space filtering and watershed clustering. In: IEEE. **Proceedings of the 2003 International Conference on Machine Learning and Cybernetics (IEEE Cat. No. 03EX693)**. [S.l.], 2003. v. 5, p. 2820–2825. Nenhuma citação no texto.

KARAR, M. E.; ALOTAIBI, B.; ALOTAIBI, M. Intelligent medical iot-enabled automated microscopic image diagnosis of acute blood cancers. **Sensors**, MDPI, v. 22, n. 6, p. 2348, 2022. Nenhuma citação no texto.

KHAN, S.; SAJJAD, M.; HUSSAIN, T.; ULLAH, A.; IMRAN, A. S. A review on traditional machine learning and deep learning models for wbcs classification in blood smear images. **IEEE Access**, IEEE, 2020. Nenhuma citação no texto.

KHANDEKAR, R.; SHASTRY, P.; JAISHANKAR, S.; FAUST, O.; SAMPATHILA, N. Automated blast cell detection for acute lymphoblastic leukemia diagnosis. **Biomedical Signal Processing and Control**, Elsevier, v. 68, p. 102690, 2021. Nenhuma citação no texto.

KHOSLA, E.; RAMESH, D. Phase classification of chronic myeloid leukemia using convolution neural networks. In: IEEE. **2018 4th International Conference on Recent Advances in Information Technology (RAIT)**. [S.l.], 2018. p. 1–6. Nenhuma citação no texto.

- KIM, Y.-j.; CHO, H. C.; CHO, H.-c. Deep learning-based computer-aided diagnosis system for gastroscopy image classification using synthetic data. **Applied Sciences**, Multidisciplinary Digital Publishing Institute, v. 11, n. 2, p. 760, 2021. Nenhuma citação no texto.
- KORNBLITH, S.; SHLENS, J.; LE, Q. V. Do better imagenet models transfer better? In: IEEE. **Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition**. [S.l.], 2019. p. 2661–2671. Nenhuma citação no texto.
- KOWSARI, K.; SALI, R.; EHSAN, L.; ADORNO, W.; ALI, A.; MOORE, S.; AMADI, B.; KELLY, P.; SYED, S.; BROWN, D. Hmic: Hierarchical medical image classification, a deep learning approach. **Information**, v. 11, n. 6, p. 318, 2020. ISSN 2078-2489. Disponível em: <<https://www.mdpi.com/2078-2489/11/6/318>>. Nenhuma citação no texto.
- KRIZHEVSKY, A.; SUTSKEVER, I.; HINTON, G. E. Imagenet classification with deep convolutional neural networks. In: **Advances in neural information processing systems**. [S.l.: s.n.], 2012. p. 1097–1105. Nenhuma citação no texto.
- KUKAČKA, J.; GOLKOV, V.; CREMERS, D. Regularization for deep learning: A taxonomy. **arXiv preprint arXiv:1710.10686**, 2017. Nenhuma citação no texto.
- LABATI, R. D.; PIURI, V.; SCOTTI, F. All-idb: The acute lymphoblastic leukemia image database for image processing. In: **IEEE International Conference on Image Processing**. [S.l.: s.n.], 2011. p. 2045–2048. Nenhuma citação no texto.
- LANDIS, J. R.; KOCH, G. G. The measurement of observer agreement for categorical data. **biometrics**, JSTOR, p. 159–174, 1977. Nenhuma citação no texto.
- LAOSAI, J.; CHAMNONGTHAI, K. Classification of acute leukemia using medical-knowledge-based morphology and cd marker. **Biomedical Signal Processing and Control**, v. 44, p. 127–137, 2018. Nenhuma citação no texto.
- LECUN, Y.; BOTTOU, L.; BENGIO, Y.; HAFFNER, P. Gradient-based learning applied to document recognition. **Proceedings of the IEEE**, IEEE, v. 86, n. 11, p. 2278–2324, 1998. Nenhuma citação no texto.
- LEE, C.-Y.; GALLAGHER, P. W.; TU, Z. Generalizing pooling functions in convolutional neural networks: Mixed, gated, and tree. In: PMLR. **Artificial intelligence and statistics**. [S.l.], 2016. p. 464–472. Nenhuma citação no texto.
- LI, L.; LIANG, J.; WENG, M.; ZHU, H. A multiple-feature reuse network to extract buildings from remote sensing imagery. **Remote Sensing**, Multidisciplinary Digital Publishing Institute, v. 10, n. 9, p. 1350, 2018. Nenhuma citação no texto.
- LI, S.; GONG, K.; LIU, C. H.; WANG, Y.; QIAO, F.; CHENG, X. Metasaug: Meta semantic augmentation for long-tailed visual recognition. In: **Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition**. [S.l.: s.n.], 2021. Nenhuma citação no texto.
- LI, Y.; ZHU, R.; MI, L.; CAO, Y.; YAO, D. Segmentation of white blood cell from acute lymphoblastic leukemia images using dual-threshold method. **Computational and mathematical methods in medicine**, Hindawi, v. 2016, n. 1, p. 1–12, 2016. Nenhuma citação no texto.

LOEY, M.; NAMAN, M.; ZAYED, H. Deep transfer learning in diagnosing leukemia in blood cells. **Computers**, Multidisciplinary Digital Publishing Institute, v. 9, n. 2, p. 29, 2020. Nenhuma citação no texto.

LYU, J.; LING, S. H. Using multi-level convolutional neural network for classification of lung nodules on ct images. In: IEEE. **2018 40th Annual International Conference of the IEEE Engineering in Medicine and Biology Society (EMBC)**. [S.l.], 2018. p. 686–689. Nenhuma citação no texto.

MADHUKAR, M.; AGAIAN, S.; CHRONOPOULOS, A. T. Deterministic model for acute myelogenous leukemia classification. In: IEEE. **IEEE International Conference on Systems, Man, and Cybernetics**. [S.l.], 2012. p. 433–438. Nenhuma citação no texto.

MISHRA, S.; MAJHI, B.; SA, P. K. A survey on automated diagnosis on the detection of leukemia: A hematological disorder. In: IEEE. **2016 3rd International Conference on Recent Advances in Information Technology (RAIT)**. [S.l.], 2016. p. 460–466. Nenhuma citação no texto.

MROZEK, K.; HEEREMA, N. A.; BLOOMFIELD, C. D. Cytogenetics in acute leukemia. **Blood reviews**, Elsevier, v. 18, n. 2, p. 115–136, 2004. Nenhuma citação no texto.

PANSOMBUT, T.; WIKAISUKSAKUL, S.; KHONGKRAPHAN, K.; PHON-ON, A. Convolutional neural networks for recognition of lymphoblast cell images. **Computational Intelligence and Neuroscience**, v. 2019, p. 1–12, 2019. Nenhuma citação no texto.

PASUPA, K.; SUNHEM, W. A comparison between shallow and deep architecture classifiers on small dataset. In: IEEE. **2016 8th International Conference on Information Technology and Electrical Engineering (ICITEE)**. [S.l.], 2016. p. 1–6. Nenhuma citação no texto.

PATEL, S. S.; WEINBERG, O. K. Diagnostic workup of acute leukemias of ambiguous lineage. **American journal of hematology**, Wiley Online Library, v. 95, n. 6, p. 718–722, 2020. Nenhuma citação no texto.

POWERS, D. M. **Evaluation: From Precision, Recall and F-Factor to ROC, Informedness, Markedness & Correlation**, School of Informatics and Engineering, Flinders University, Adelaide, Australia. [S.l.], 2007. Nenhuma citação no texto.

PUTTAGUNTA, M.; RAVI, S. Medical image analysis based on deep learning approach. **Multimedia Tools and Applications**, v. 80, n. 16, p. 24365–24398, 2021. Nenhuma citação no texto.

RASTOGI, P.; KHANNA, K.; SINGH, V. Leufeatx: Deep learning–based feature extractor for the diagnosis of acute leukemia from microscopic images of peripheral blood smear. **Computers in Biology and Medicine**, Elsevier, p. 105236, 2022. Nenhuma citação no texto.

RAWAT, J.; SINGH, A.; HS, B.; VIRMANI, J.; DEVGUN, J. S. Computer assisted classification framework for prediction of acute lymphoblastic and acute myeloblastic leukemia. **Biocybernetics and Biomedical Engineering**, v. 37, n. 4, p. 637 – 654, 2017. Nenhuma citação no texto.

REZATOFIGHI, S. H.; SOLTANIAN-ZADEH, H. Automatic recognition of five types of white blood cells in peripheral blood. **Computerized Medical Imaging and Graphics**, v. 35, n. 4, p. 333 – 343, 2011. ISSN 0895-6111. Nenhuma citação no texto.

RODRIGUES, L. F.; BACKES, A. R.; TRAVENÇOLO, B. A. N.; OLIVEIRA, G. M. B. de. Optimizing a deep residual neural network with genetic algorithm for acute lymphoblastic leukemia classification. **Journal of Digital Imaging**, Springer, p. 1–15, 2022. Nenhuma citação no texto.

ROLLINS-RAVAL, M.; RAVAL, J.; CONTIS, L. Experience with cellavision dm96 for peripheral blood differentials in a large multi-center academic hospital system. **Journal of Pathology Informatics**, v. 3, n. 29, p. 1–9, 2012. Nenhuma citação no texto.

ROSENFELD, G. H.; FITZPATRICK-LINS, K. A coefficient of agreement as a measure of thematic classification accuracy. **Photogrammetric engineering and remote sensing**, v. 52, n. 2, p. 223–227, 1986. Nenhuma citação no texto.

RUSSAKOVSKY, O.; DENG, J.; SU, H.; KRAUSE, J.; SATHEESH, S.; MA, S.; HUANG, Z.; KARPATY, A.; KHOSLA, A.; BERNSTEIN, M. et al. Imagenet large scale visual recognition challenge. **International journal of computer vision**, Springer, v. 115, n. 3, p. 211–252, 2015. Nenhuma citação no texto.

SAEED, A.; SHOUKAT, S.; SHEHZAD, K.; AHMAD, I.; ESHMAWI, A.; AMIN, A. H.; TAG-ELDIN, E. et al. A deep learning-based approach for the diagnosis of acute lymphoblastic leukemia. **Electronics**, Multidisciplinary Digital Publishing Institute, v. 11, n. 19, p. 3168, 2022. Nenhuma citação no texto.

SARRAFZADEH, O.; DEHNAVI, A. M. Nucleus and cytoplasm segmentation in microscopic images using k means clustering and region growing. **Advanced Biomedical Research**, p. 79–87, December 2015. Nenhuma citação no texto.

SARRAFZADEH, O.; RABBANI, H.; DEHNAVI, A. M.; TALEBI, A. Detecting different subtypes of acute myelogenous leukemia using dictionary learning and sparse representation. In: IEEE. **2015 IEEE International Conference on Image Processing (ICIP)**. [S.l.], 2015. p. 3339–3343. Nenhuma citação no texto.

SARRAFZADEH, O.; RABBANI, H.; DEHNAVI, A. M.; TALEBI, A. Detecting different subtypes of acute myelogenous leukemia using dictionary learning and sparse representation. In: **IEEE International Conference on Image Processing**. [S.l.]: IEEE, 2015. p. 3339–3343. Nenhuma citação no texto.

SARRAFZADEH, O.; RABBANI, H.; TALEBI, A.; BANAEM, H. U. Selection of the best features for leukocytes classification in blood smear microscopic images. In: **SPIE Medical Imaging**. [S.l.: s.n.], 2014. v. 9041. Nenhuma citação no texto.

SASHANK, G. V. S.; JAIN, C.; VENKATESWARAN, N. Detection of acute lymphoblastic leukemia by utilizing deep learning methods. In: **Machine Vision and Augmented Intelligence—Theory and Applications**. [S.l.]: Springer, 2021. p. 453–467. Nenhuma citação no texto.

SAWYERS, C. L.; DENNY, C. T.; WITTE, O. N. Leukemia and the disruption of normal hematopoiesis. **Cell**, Elsevier, v. 64, n. 2, p. 337–350, 1991. Nenhuma citação no texto.

- SHAFIQUE, S.; TEHSIN, S. Acute lymphoblastic leukemia detection and classification of its subtypes using pretrained deep convolutional neural networks. **Technology in Cancer Research and Treatment**, v. 17, p. 1–7, september 2018. Nenhuma citação no texto.
- SHORTEN, C.; KHOSHGOFTAAR, T. M. A survey on image data augmentation for deep learning. **Journal of Big Data**, Springer, v. 6, n. 1, p. 60, 2019. Nenhuma citação no texto.
- SIMONYAN, K.; ZISSERMAN, A. Very deep convolutional networks for large-scale image recognition. **International Conference on Learning Representations (ICLR)**, 2014. Nenhuma citação no texto.
- SINGHAL, V.; SINGH, P. Texture features for the detection of acute lymphoblastic leukemia. In: _____. **Proceedings of International Conference on ICT for Sustainable**. Singapore: Springer Singapore, 2016. p. 535–543. Nenhuma citação no texto.
- Sipes, R.; Li, D. Using convolutional neural networks for automated fine grained image classification of acute lymphoblastic leukemia. In: **Proceedings of the 3rd International Conference on Computational Intelligence and Applications**. [S.l.: s.n.], 2018. p. 157–161. Nenhuma citação no texto.
- SOERJOMATARAM, I.; BRAY, F. Planning for tomorrow: global cancer incidence and the role of prevention 2020–2070. **Nature Reviews Clinical Oncology**, Nature Publishing Group, p. 1–10, 2021. Nenhuma citação no texto.
- SOLLICH, P.; KROGH, A. Learning with ensembles: How overfitting can be useful. **Advances in neural information processing systems**, v. 8, p. 190–196, 1995. Nenhuma citação no texto.
- SOUZA, L. M. d.; GORINI, M. I. P. C. Diagnósticos de enfermagem em adultos com leucemia mieloide aguda. **Revista gaúcha de enfermagem. Porto Alegre. Vol. 27, n. 3 (set. 2006), p. 417-425**, 2006. Nenhuma citação no texto.
- SRINIVASAN, C.; DUBEY, S.; GANESHBABU, T. Complex texture features for glaucomatous image classification system using fundus images. v. 2, n. 13, 2016. Nenhuma citação no texto.
- SZEGEDY, C.; LIU, W.; JIA, Y.; SERMANET, P.; REED, S.; ANGUELOV, D.; ERHAN, D.; VANHOUCHE, V.; RABINOVICH, A. Going deeper with convolutions. In: **Proceedings of the IEEE conference on computer vision and pattern recognition**. [S.l.: s.n.], 2015. p. 1–9. Nenhuma citação no texto.
- SZEGEDY, C.; VANHOUCHE, V.; IOFFE, S.; SHLENS, J.; WOJNA, Z. Rethinking the inception architecture for computer vision. In: **Proceedings of the IEEE conference on computer vision and pattern recognition**. [S.l.: s.n.], 2016. p. 2818–2826. Nenhuma citação no texto.
- THANH, T.; VUNUNU, C.; ATOEV, S.; LEE, S.-H.; KWON, K.-R. Leukemia blood cell image classification using convolutional neural network. **International Journal of Computer Theory and Engineering**, v. 10, n. 2, p. 54–58, 2018. Nenhuma citação no texto.
- TRAN, T.; PARK, J.-H.; KWON, O.-H.; MOON, K.-S.; LEE, S.-H.; KWON, K.-R. Classification of leukemia disease in peripheral blood cell images using convolutional neural network. **Journal of Korea Multimedia Society**, Korea Multimedia Society, v. 21, n. 10, p. 1150–1161, 2018. Nenhuma citação no texto.

TRAVLOS, G. S. Normal structure, function, and histology of the bone marrow. **Toxicologic Pathology**, v. 34, n. 5, p. 548–565, 2006. Nenhuma citação no texto.

ULLAH, M. Z.; ZHENG, Y.; SONG, J.; ASLAM, S.; XU, C.; KIAZOLU, G. D.; WANG, L. An attention-based convolutional neural network for acute lymphoblastic leukemia classification. **Applied Sciences**, Multidisciplinary Digital Publishing Institute, v. 11, n. 22, p. 10662, 2021. Nenhuma citação no texto.

UPRETI, M.; PANDEY, C.; BIST, A. S.; RAWAT, B.; HARDINI, M. Convolutional neural networks in medical image understanding. **Aptisi Transactions on Technopreneurship**, v. 3, n. 2, p. 6–12, 2021. Nenhuma citação no texto.

VALE, A. M. P. G.; GUERREIRO, A. M. G.; NETO, A. D. D.; JUNIOR, G. B. C.; LEITÃO, V. C. L. T. de S.; MARTINS, A. M. Automatic segmentation and classification of blood components in microscopic images using a fuzzy approach. **Revista Brasileira de Engenharia Biomédica**, scielo, v. 30, p. 341–354, 2014. ISSN 1517-3151. Nenhuma citação no texto.

VOGADO, L.; VERAS, R.; AIRES, K.; ARAÚJO, F.; SILVA, R.; PONTI, M.; TAVARES, J. M. R. Diagnosis of leukaemia in blood slides based on a fine-tuned and highly generalisable deep learning model. **Sensors**, Multidisciplinary Digital Publishing Institute, v. 21, n. 9, p. 2989, 2021. Nenhuma citação no texto.

VOGADO, L. H.; VERAS, R. de M.; ANDRADE, A. R.; SILVA, R. R. e; ARAUJO, F. H. D.; MEDEIROS, F. N. D. Unsupervised leukemia cells segmentation based on multi-space color channels. In: IEEE. **2016 IEEE International Symposium on Multimedia (ISM)**. [S.l.], 2016. p. 451–456. Nenhuma citação no texto.

VOGADO, L. H. S.; VERAS, R. M. S.; ARAÚJO, F. H. D.; SILVA, R. R. V. e; AIRES, K. R. T. Leukemia diagnosis in blood slides using transfer learning in cnns and SVM for classification. **Engineering Applications of Artificial Intelligence**, v. 72, p. 415–422, 2018. Nenhuma citação no texto.

WALT, S. Van der; SCHÖNBERGER, J. L.; NUNEZ-IGLESIAS, J.; BOULOGNE, F.; WARNER, J. D.; YAGER, N.; GOUILLART, E.; YU, T. scikit-image: image processing in python. **PeerJ**, PeerJ Inc., v. 2, p. e453, 2014. Nenhuma citação no texto.

WANG, Y.; HUANG, G.; SONG, S.; PAN, X.; XIA, Y.; WU, C. **Regularizing Deep Networks with Semantic Data Augmentation**. 2021. Nenhuma citação no texto.

XIE, J.; MA, Z.; CHANG, D.; ZHANG, G.; GUO, J. Gpca: A probabilistic framework for gaussian process embedded channel attention. **IEEE Transactions on Pattern Analysis and Machine Intelligence**, n. 1, p. 1–1, 2021. Nenhuma citação no texto.

ZHAO, W.; JIANG, W.; QIU, X. Deep learning for covid-19 detection based on ct images. **Scientific Reports**, Nature Publishing Group, v. 11, n. 1, p. 1–12, 2021. Nenhuma citação no texto.

ZHENG, X.; WANG, Y.; WANG, G.; CHEN, Z. Fast and robust segmentation of white blood cell images by self-supervised learning. **Micron**, v. 107, p. 55–71, 2018. Nenhuma citação no texto.