

Explorando Arquiteturas de Redes Neurais Profundas na Classificação de Imagens de Cariótipos Humanos

Francisco das C. Imperes Filho¹, Vinicius P. Machado¹, Semiramis J. H. do Monte², Alan R. F. dos Santos¹, Leonardo P. de Sousa¹, Adalberto S. da Silva², Ester M. Pereira³ and Rodrigo de M. S. Veras¹

¹Núcleo de Computação de Alto Desempenho – UFPI – Teresina – PI – Brazil

²Laboratório de Imunogenética e Biologia Molecular (UFPI)

³Departamento de Clínica Geral (CCS/UFPI)

{fcoimperes, vinicius, semiramis, arfs,
leonardosousa}@ufpi.edu.br

adalberto@ufpi.edu.br, estermperreira@gmail.com

rveras@ufpi.edu.br

Abstract. *Chromosome analysis, a clinically crucial practice traditionally performed by geneticists, can be susceptible to fatigue over time, affecting the quality of diagnoses. This paper explored the automated classification of chromosome images using various deep-learning architectures. We evaluated 23 pairs of human chromosomes in a multiclass task, revealing promising results. The superior performances of the DenseNet169 architecture are highlighted, achieving an accuracy, precision, recall, and F1-Score of 98.77%. The Kappa agreement index reached an "Excellent" level (0.99), while a low standard deviation (0.002) highlighted the consistency of the metrics, providing reliability and predictability to the proposed model.*

Resumo. *A análise cromossômica, uma prática clinicamente crucial realizada tradicionalmente por geneticistas, pode ser suscetível à fadiga ao longo do tempo, afetando a qualidade dos diagnósticos. Neste artigo, exploramos a classificação automatizada de imagens de cromossomos por meio de diversas arquiteturas de redes neurais profundas. Avaliamos 23 pares de cromossomos humanos em uma tarefa multiclasse, revelando resultados promissores. Destacam-se os desempenhos superiores da arquitetura DenseNet169, alcançando uma acurácia, precisão, recall e F1-Score de 98,77%. O índice de concordância Kappa atingiu um nível "Excelente" (0,99), enquanto um baixo desvio padrão (0,002) ressaltou a consistência das métricas, conferindo confiabilidade e previsibilidade ao modelo proposto.*

1. Introdução

A cariotipagem é fundamental para a triagem de doenças genéticas em um estágio inicial [Anh et al. 2022]. Este método analisa células de um indivíduo para determinar seu padrão cromossômico [Abid and Hamami 2018]. Basicamente ela corresponde em alinhar e parear os cromossomos pelo seu comprimento para realizar a análise mais

rapidamente. Conforme demonstra a Figura 1, a técnica consiste na microfotografia, em sequência, dos pares de cromossomos e permite identificar um indivíduo normal ou com alguma alteração genética. Um protocolo muito utilizado para realizar a cariotipagem é a coloração dos cromossomos em bandas (faixas), denominado de bandeamento. Banda G, banda C, banda R e banda Q são os métodos de coloração mais comuns.

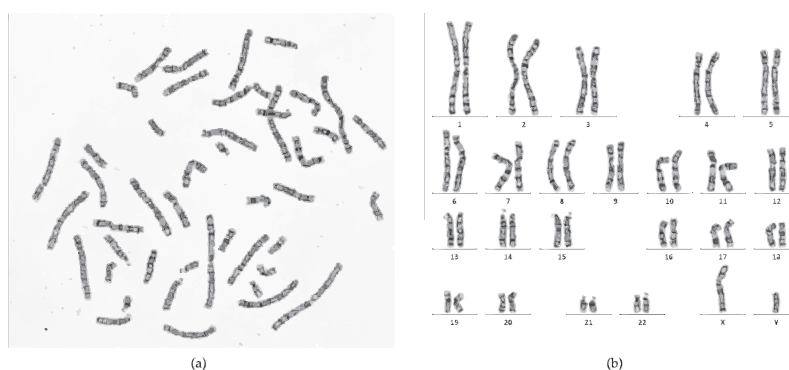


Figura 1. Análise de cariotipagem cromossômica: (a) microfotografia de cromossomos corados com banda G, e (b), cariótipo cromossômico correspondente. Fonte: [Lin et al. 2022].

Nesse contexto, a utilização de métodos computacionais têm sido utilizados para prover modelos de classificação de cromossomos para auxiliar geneticistas no processo de análises clínicas. Dentre estes métodos, podemos citar pesquisas que utilizam *Convolutional Neural Networks* (CNN), para propor modelos confiáveis na perspectiva de agilizar o processo manual relacionado à classificação de cromossomos.

Esses métodos têm o potencial de agilizar a identificação de anomalias cromossômicas, permitindo uma avaliação mais detalhada e confiável para o diagnóstico preciso de doenças genéticas. Este artigo investigou e avaliou diversas arquiteturas de CNN na classificação de imagens de cromossomos humanos. Nosso objetivo foi analisar a capacidade dessas arquiteturas em identificar e categorizar com precisão as variações cromossômicas em cariótipos, visando avanços na genética e diagnósticos médicos.

Além disso, o artigo inclui a combinação das previsões de múltiplas CNNs para gerar uma única previsão final. Esse método, chamado Aprendizado de Conjunto ou Comitê, visa aproveitar as diversas perspectivas fornecidas por cada modelo individual, resultando em previsões mais precisas e robustas.

Este artigo está organizado da seguinte forma: a Seção 2 apresenta os trabalhos relacionados. Na Seção 3, é detalhada a metodologia proposta. Os resultados são apresentados na Seção 4, seguidos por discussões na Seção 5 e conclusões na Seção 6.

2. Trabalhos Relacionados

Nos últimos anos, várias metodologias foram desenvolvidas para classificar imagens de cromossomos. Apresentaremos a seguir aquelas que se destacaram pelos melhores resultados obtidos.

Anh et al. [Anh et al. 2022] propuseram uma arquitetura de rede neural convolucional para classificação de tipo e polaridade da imagem cromossômica, ou

seja, *Efficient Type and Polarity Classification* (ETPC), que é alavancada a partir do desenvolvimento da família EfficientNet. Na proposta relatam que o classificador ETPC com função de perda de classificação ponderada foi projetado para eficiência no processo de treinamento. Os pesquisadores realizaram o experimento com dois cenários de treinamento em quatro conjuntos de dados clínicos. O experimento em um conjunto de dados de 28.225 imagens cromossômicas originais demonstraram que a rede proposta alcançou resultados comparáveis, com uma acurácia de 95,3% para a classificação de tipo e 99% para a classificação de polaridade, tendo um número de parâmetros significativamente menor do que métodos de última geração.

O trabalho proposto por [Lin et al. 2022] teve como objetivo construir e reproduzir, avaliar e comparar modelos e algoritmos de classificação de cromossomos existentes em um conjunto de dados de classificação de cromossomos em nível de mais de 120.000 amostras. Segundo os autores o conjunto de dados permite o *benchmarking* (avaliação comparativa) e a construção de linhas de base de classificação cromossômica adequada para diferentes cenários. O conjunto de dados clínicos consiste em 126.453 instâncias de cromossomos de Banda-G preservando a privacidade de 2.763 cariótipos de 408 indivíduos. Os pesquisados ressaltam que a melhor linha de base com 0,99 de precisão, *recall* e *F1-score* e com 99,33% de acurácia, relata desempenho de classificação de última geração.

No estudo de [Saranya and Lakshmi 2023] foi aplicada uma metodologia que utiliza Redes Neurais Convolucionais para a classificação de cromossomos e a detecção de anormalidades cromossômicas, com o objetivo de superar as limitações da cariotipagem manual. Utilizando um conjunto de dados de 14.144 imagens de cromossomos, o estudo aplica técnicas de pré-processamento e análise de imagens, incluindo a Transformada Discreta de Wavelet e arquiteturas CNN como VGG-16, ResNet50, e Inception, para realizar a classificação. O trabalho alcançou acurácia de 92,5% na classificação e introduziu uma nova métrica de avaliação, "percentagem de cariótipos bem classificados", com a qual atingiu acurácia de 91,3%. Dentro do contexto do trabalho, não foi possível identificar a disponibilidade da base de dados utilizada.

O estudo de [Xia et al. 2023], intitulado *KaryoNet*, introduziu uma metodologia para o reconhecimento de cromossomos, focando na identificação de interações de longo alcance entre cromossomos em células metafásicas¹ para aprimorar a classificação de tipos e polaridades cromossômicas em cariótipos normais e anormalidades numéricas. Empregando uma abordagem *end-to-end* que incluiu técnicas de extração de características via CNN, *Masked Feature Interaction Module* (MFIM), *Deep Assignment Module* (DAM), *ResNet50 backbone*, *Transformer encoder*, *Bi-RNN blocks* e *Gated Recurrent Unit* (GRU), o *KaryoNet* alcançou acurácias notáveis de 98,41% para cromossomos *R-band* e 99,58% para cromossomos *G-band* em avaliações clínicas, superando o *software* comercial *Ikaros* com uma acurácia média de 99,19% em comparação aos 90,17% registrados por este último.

3. Materiais e Métodos

Na análise comparativa das classificações de imagens de cromossomos utilizamos as Redes Neurais Convolucionais pré-treinadas *DenseNet121*, *DenseNet169*,

¹Células metafásicas são um estágio específico no ciclo celular, ocorrendo após a prófase.

DenseNet201 [Huang et al. 2017], *EfficientNetV2B3* [Tan and Le 2021], *InceptionV3* [Szegedy et al. 2016], *ResNet50* [He et al. 2016], *VGG16*, *VGG19* [Simonyan and Zisserman 2014] e *Xception* [Chollet 2017]. O critério para a escolha destas CNNs foi devido à maioria dos trabalhos relacionados neste artigo utilizarem essas arquiteturas, facilitando a comparação com o estado da arte.

As arquiteturas CNN foram treinadas utilizando o método estatístico *K-fold Cross-Validation* em 5 épocas, 5 passos por época, sendo aplicadas para cada arquitetura e classificação as estratégias *Shallow Fine-Tuning* (SFT) e *Deeply Fine-Tuning* (DFT) [Izadyazdanabadi et al. 2018]. Este trabalho destaca os resultados da estratégia DFT por se mostrar superior à SFT, independentemente da arquitetura utilizada.

Devido a limitações de recursos computacionais, o tamanho do lote de treinamento foi definido como 8, resultando em 2.256 iterações para as estratégias mencionadas. Utilizamos a função *categorical cross-entropy* para treinar as CNNs em classificação multiclasse, empregando os otimizadores *ADAM* e *SGD* para os experimentos.

3.1. Base de Imagens

O conjunto de dados proposto por [Lin et al. 2022] foi usado como referencial para realizar a análise comparativa entre as CNNs. O conjunto de dados clínicos contém 126.453 imagens onde 125.219 apresentam amostras individuais de cromossomos corados em Banda G. Elas foram coletadas do *Guangdong Women and Children Hospital* e segmentadas de 2.763 cariótipos de cromossomos de 225 indivíduos.

O repositório contém referências à classificação clínica de cromossomos com 24 categorias, incluindo 22 categorias de autossomos marcados de 0-21 e cromossomos sexuais X, Y marcados como 22 e 23, respectivamente. Todas as imagens têm resolução de 300×300 pixels. A Figura 2 representam algumas imagens levando em consideração as categorias a qual os cromossomos pertencem.



Figura 2. Cromossomos Categorias 1 (a) e 15 (b). Fonte: [Lin et al. 2022].

A Figura 2(a) refere-se a cromossomos Categoria 1 e a 2(b) de Categoria 15. Com base nas imagens exibidas na Figura 2 observa-se que mesmo pertencendo à mesma categoria os cromossomos diferem entre si na forma como foram capturados e, conseqüentemente, na maneira como devem ser analisados pelo geneticista. A disposição dos cromossomos nas imagens é um grande desafio para o processo de classificação de cromossomos, seja ela manual ou utilizando recursos computacionais.

3.2. Aumento de Dados

O treinamento de CNNs é computacionalmente custoso devido ao grande número de parâmetros e à necessidade de uma grande quantidade de dados. Para lidar com isso, muitos estudos utilizam aumento de dados, que envolve a criação de novas imagens

por meio de transformações, como rotação e inversão. Essa técnica ajuda a reduzir o overfitting e a melhorar a capacidade de generalização da rede [Perez and Wang 2017].

Neste trabalho, durante a etapa de refinamento das CNNs, foi utilizado o aumento de dados. A técnica consistiu na geração de novas amostras durante o treinamento, aplicando operações aleatórias em *batches*. Seguem os procedimentos aplicados com seus respectivos parâmetros: centralização por recurso (`featurewise_center=True`), normalização por recurso (`featurewise_std_normalization=True`), deslocamento Horizontal (`width_shift_range=0.1`), deslocamento vertical (`height_shift_range=0.08`), rotação (`rotation_range=4`), espelhamento horizontal (`horizontal_flip=True`), amplitude de zoom (`zoom_range=0.02`), função de pré-processamento (`preprocessing_function=add_noise`) e modo de preenchimento (`fill_mode="constant"`). Por fim, foi realizada a normalização dos pixels das imagens de entrada para 0 e 1 (`rescale=1./255`).

3.3. Arquiteturas de CNNs

Foram exploradas e avaliadas as arquiteturas CNNs projetadas para o *ImageNet Large Scale Visual Recognition Challenge* (ILSVRC) [Russakovsky et al. 2015]. De acordo com [Kornblith et al. 2019], o desempenho da arquitetura no conjunto de dados *ImageNet* é um indicativo de sua capacidade de transferência para outros conjuntos de dados de imagens naturais. Além disso, o desempenho em estudos da literatura, como o de Vogado et al. (2021) [Vogado et al. 2021], foi outro critério considerado na seleção das arquiteturas no estado da arte.

As redes neurais sequenciais, como *VGG-16* e *VGG-19*, oferecem flexibilidade para ajustes na arquitetura, enquanto as redes residuais e *Inception* demonstraram desempenho superior no desafio ILSVRC [Russakovsky et al. 2015]. Comparativamente, *ResNet50*, *InceptionV3* e as variantes *DenseNet* (*DenseNet121*, *DenseNet169* e *DenseNet201*) possuem menos parâmetros em comparação com as *VGGNets*, mas apresentam maior profundidade, indicada pelo número de filtros nas operações de convolução. A arquitetura *Xception*, uma evolução da *Inception* [Szegedy et al. 2016], supera marginalmente a *InceptionV3* no conjunto de dados *ImageNet*. Por fim, a rede *EfficientNetV2B3* é reconhecida por sua eficiência em termos de velocidade de treinamento e uso otimizado de parâmetros em comparação com modelos anteriores da mesma família.

3.4. Transferência de Aprendizagem e Ajuste Fino

Transferência de Aprendizagem é uma técnica utilizada em aprendizagem profunda para beneficia-se de pesos pré-treinados em grandes volumes de dados. Dois métodos são frequentemente empregadas quando os pesos pré-treinados são usados para transferência de aprendizado.

O primeiro envolve a extração de características utilizando as camadas totalmente conectadas da CNN como entrada para classificadores do estado da arte. O segundo é conhecida como ajuste fino ou *fine-tuning*, onde os pesos das camadas convolucionais são congelados e apenas as camadas totalmente conectadas são reajustadas com uma baixa taxa de aprendizado [Santos et al. 2021]. Embora mais eficaz, o ajuste fino ainda requer poder computacional significativo, já que envolve retreinar a CNN com um novo conjunto de dados, adaptando o modelo para o novo domínio desejado.

Existem duas estratégias de ajuste fino: *Shallow Fine-Tuning* (SFT) e *Deeply Fine-Tuning* (DFT) [Izadyyazdanabadi et al. 2018]. A SFT consiste no congelamento das camadas iniciais da CNN, geralmente os conjuntos de camadas convolucionais. Essas camadas são consideradas mais gerais e conseguem representar características de forma, textura e cor. As camadas superiores costumam ser específicas de domínio, carregando conteúdo semântico dos rótulos das instâncias. A DFT permite treinar toda a rede, adaptando até as primeiras camadas, exigindo um custo computacional mais alto, assim como maior quantidade de dados. Esta proposta enfatiza a utilização da técnica *fine-tuning* DFT.

3.5. Métricas de Avaliação e Ambiente de Experimentos

A avaliação das CNNs foi conduzida com base nos valores obtidos pela matriz de confusão. A partir dessa matriz, são derivados quatro valores: Verdadeiro Positivo (VP), Falso Positivo (FP), Falso Negativo (FN) e Verdadeiro Negativo (VN). Nesse contexto, a avaliação dos experimentos realizados nesta pesquisa é calculada levando em consideração as seguintes métricas: Acurácia (A), Precisão (P), *Recall* (R), *F1-score* e o Coeficiente *Kappa* (κ). O Desvio Padrão (DP) foi empregado para indicar a variabilidade dos valores no conjunto de dados utilizado como referência para a análise comparativa entre as CNNs [Sokolova and Lapalme 2009]. Além disso, a explicabilidade (ou interpretabilidade) é outra técnica relevante adotada [Hsu and Li 2023]. Também consideramos a utilização da estratégia de Aprendizagem de Conjunto, também conhecida como Comitê ou *Ensemble Learning* em inglês, para combinar as previsões das CNNs que apresentaram os melhores resultados nos experimentos, a fim de gerar uma única previsão final [Moon et al. 2020].

3.6. Configuração do ambiente dos experimentos

Todos os experimentos foram realizados utilizando-se da Linguagem de Programação (LP) *Python* 3.9.13, em conjunto com a plataforma *Machine Learning TensorFlow*, rodando em um processador Intel® i9 12^a geração, 64GB de memória RAM e uma GPU NVIDIA GeForce RTX 3080 com 10240MiB de memória.

4. Resultados

O propósito desta seção é expor e interpretar os resultados dos experimentos realizados neste trabalho. A discussão é orientada no sentido de entender como as diferentes abordagens influenciaram no desempenho do modelo, utilizando o conjunto de dados ChromosomeNet [Lin et al. 2022].

4.1. Classificação Multiclasse: Estratégia DFT com otimizador ADAM.

A Tabela 1 exibe os resultados dos experimentos conduzidos com as arquiteturas de redes CNN pré-treinadas, utilizando a estratégia DFT com o otimizador ADAM.

A Tabela 1 expressa que as arquiteturas *DenseNet169* e *Xception* obtiveram os melhores desempenhos com resultados em pontos percentuais bem próximos uma da outra. A primeira teve desempenho levemente superior atingindo acurácia, *recall* e *F1-Score* de 98,77%, e precisão de 98,78%. O coeficiente *Kappa* foi de 0,99, indicando uma concordância quase perfeita, e o desvio padrão das métricas foi de 0,002 para as

Tabela 1. Resultados dos experimentos utilizando o conjunto de dados ChromosomeNet com 125.219 imagens. Estratégia DFT com otimizador ADAM (melhores resultados destacados em negrito).

Arquitetura	A (%)	P (%)	R (%)	F1-Score (%)	K
DenseNet121	98,74 ± 0,002	98,75 ± 0,002	98,74 ± 0,002	98,74 ± 0,002	0,99 ± 0,002
DenseNet169	98,77 ± 0,002	98,78 ± 0,002	98,77 ± 0,002	98,77 ± 0,002	0,99 ± 0,002
DenseNet201	98,42 ± 0,005	98,46 ± 0,005	98,42 ± 0,005	98,42 ± 0,005	0,98 ± 0,005
EfficientNetV2B3	98,74 ± 0,001	98,75 ± 0,001	98,74 ± 0,001	98,74 ± 0,001	0,99 ± 0,001
InceptionV3	98,75 ± 0,002	98,76 ± 0,002	98,75 ± 0,002	98,75 ± 0,002	0,99 ± 0,002
Resnet50	98,26 ± 0,005	98,28 ± 0,005	98,26 ± 0,005	98,26 ± 0,005	0,98 ± 0,006
VGG-16	98,09 ± 0,005	98,12 ± 0,005	98,09 ± 0,005	98,09 ± 0,005	0,98 ± 0,006
VGG-19	98,09 ± 0,003	98,11 ± 0,003	98,09 ± 0,003	98,09 ± 0,003	0,98 ± 0,003
Xception	98,76 ± 0,002	98,77 ± 0,002	98,76 ± 0,002	98,76 ± 0,002	0,99 ± 0,002

duas arquiteturas, demonstrando alta estabilidade nos resultados. De modo diferente, as arquiteturas *VGG-16* e *VGG-19* obtiveram desempenhos inferiores em termos de acurácia, registrando 98,06%. No entanto, é importante observar que, mesmo com acurácia mais baixa, essas arquiteturas mantiveram um alto coeficiente *Kappa* de 0,98 e desvio padrão variando entre 0,003 e 0,006, indicando a eficácia geral da técnica DFT.

4.2. Classificação Multiclasse: Estratégia DFT com otimizador SGD.

Esta sub-seção apresenta os resultados da classificação utilizando as configurações das arquiteturas de redes CNN pré-treinadas. Os resultados dos experimentos com o uso do otimizador SGD em conjunto com a estratégia DFT são exibidos na Tabela 2.

Tabela 2. Resultados dos experimentos utilizando o conjunto de dados ChromosomeNet com 125.219 imagens. Estratégia DFT com otimizador SGD (melhores resultados destacados em negrito).

Arquitetura	A (%)	P (%)	R (%)	F1-Score (%)	K
DenseNet121	97,56 ± 0,008	97,57 ± 0,008	97,56 ± 0,008	97,56 ± 0,008	0,97 ± 0,008
DenseNet169	97,90 ± 0,006	97,92 ± 0,006	97,90 ± 0,006	97,91 ± 0,006	0,99 ± 0,006
DenseNet201	97,97 ± 0,007	97,99 ± 0,007	97,97 ± 0,007	97,97 ± 0,007	0,98 ± 0,007
EfficientNetV2B3	87,70 ± 0,107	87,79 ± 0,104	87,70 ± 0,107	87,26 ± 0,113	0,87 ± 0,112
InceptionV3	97,37 ± 0,007	97,39 ± 0,007	97,37 ± 0,007	97,38 ± 0,007	0,97 ± 0,007
VGG16	97,55 ± 0,002	97,58 ± 0,002	97,55 ± 0,002	97,56 ± 0,002	0,97 ± 0,002
VGG19	97,63 ± 0,004	97,65 ± 0,004	97,63 ± 0,004	97,63 ± 0,004	0,98 ± 0,004
Xception	90,41 ± 0,047	91,10 ± 0,050	90,41 ± 0,047	90,44 ± 0,047	0,90 ± 0,049

A Tabela 2 expressa que a arquitetura *DenseNet1201* obteve o melhor desempenho atingindo acurácia, *recall* e *F1-Score* de 97,97%, e precisão de 97,99%. O coeficiente *Kappa* foi de 0,98, indicando uma concordância quase perfeita, e o desvio padrão das métricas foi de 0,007, demonstrando alta estabilidade nos resultados. Por outro lado, a arquitetura *EfficientNetV2B3* alcançou desempenho inferior em termos de acurácia, registrando 87,70%. O coeficiente *Kappa* de 0,87 indica fator de eficácia para esta métrica, destacando a eficiência da estratégia DFT.

5. Discussão

Com base nas análises realizadas dos experimentos, observamos uma variedade de resultados influenciados pela escolha da arquitetura da rede CNN e do otimizador

utilizado (ADAM ou SGD).

Independentemente do otimizador utilizado, todas as arquiteturas testadas alcançaram acurácias acima de 90%. Com o otimizador ADAM, arquiteturas como *DenseNet169*, *Xception*, *Inception*, *DenseNet121* e *EfficientNetV2B3*, frequentemente emergiram como as mais eficazes, alcançando acurácias superiores a 98%. Com o otimizador SGD, a arquitetura *DenseNet201* foi particularmente superior, exibindo a maior acurácia (97,76%), precisão, *F1-Score*, e coeficiente *Kappa*, além de um baixo desvio padrão, indicando resultados confiáveis e consistentes.

Com relação ao impacto do otimizador (ADAM vs. SGD) na performance dos modelos, enquanto o otimizador SGD proporcionou resultados robustos e consistentes, foi notável que a utilização do otimizador ADAM, especialmente em conjunto com a técnica DFT, resultou em um desempenho mais alto. Isso foi evidenciado por acurácias superiores alcançadas pelas arquiteturas em várias instâncias, como a *DenseNet169*, que obteve os melhores resultados dentre todos os experimentos realizados. A escolha do otimizador ADAM pareceu contribuir significativamente para a eficácia do modelo, promovendo uma maior precisão e estabilidade nos resultados.

5.1. Desempenho do Modelo

Ao analisar os resultados dos exemplos nesta subseção, foi escolhida a arquitetura *DenseNet169* devido ao seu desempenho superior em comparação com as outras listadas na Tabela 1. A Figura 3 mostram imagens de cromossomos das classes 1, 9 e 1, respectivamente, com seus Mapas de Saliência (MS). Todos eles destacam as áreas mais influentes na decisão da rede neural. Esses mapas são calculados usando gradientes para entender como variações sutis nos pixels da imagem afetam a saída da rede. As regiões mais iluminadas indicam quais partes da imagem foram mais determinantes para a previsão da rede neural, evidenciando a sensibilidade da rede a essas áreas específicas da imagem.

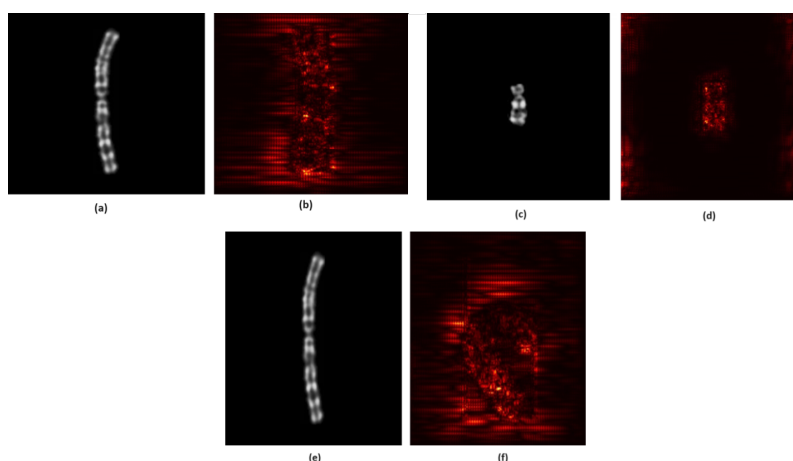


Figura 3. Imagens de cromossomos e seus respectivos mapas de ativação.

Na Figura 3(b), o MS destaca regiões fora da forma regular de um cromossomo da classe 1; no entanto, a rede neural o classificou corretamente. Por outro lado, a Figura 3(d) ilustra um caso em que a rede neural identificou corretamente as áreas destacadas no MS

como correspondentes a um cromossomo da classe 9, classificando-o de forma precisa. Em contraste, na Figura 3(f), que também representa um cromossomo da classe 1, são destacadas regiões específicas da imagem que não se assemelham a um cromossomo da classe 1, resultando em uma classificação final incorreta.

A Figura 4 ilustra uma análise detalhada do desempenho do modelo proposto neste trabalho. A Subfigura 4(a) foca na evolução da acurácia, apresentando-a para os conjuntos de treinamento e validação. As linhas representam as médias da métrica, enquanto as áreas sombreadas demonstram os desvios padrão, expondo as flutuações ao longo das épocas. Analogamente, a Subfigura 4(b) explora a progressão da perda (*Loss*) durante o treinamento, mostrando-a para os conjuntos de treinamento e validação. As linhas indicam a média da perda ao longo das épocas, e as áreas sombreadas refletem os desvios padrão, enfatizando as variações. A diminuição progressiva desses valores aponta para uma adaptação apropriada do modelo aos dados, sugerindo a inexistência de *overfitting* e um ajuste eficaz às informações de validação. Essas análises visuais fornecem *insights* essenciais sobre a estabilidade e eficácia do modelo durante o processo de treinamento.

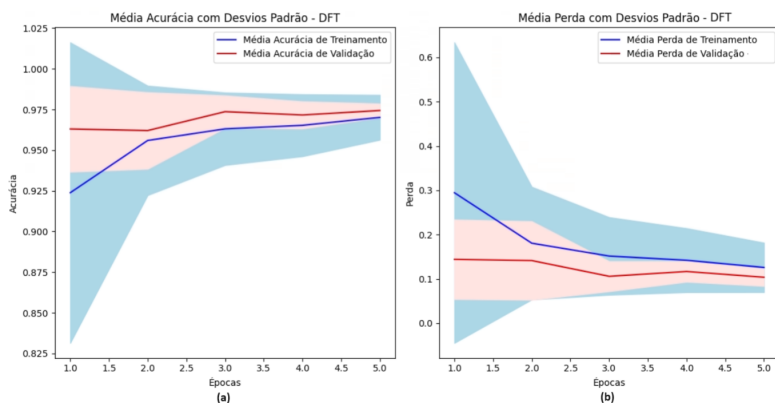


Figura 4. Desempenho do modelo: (a) evolução da acurácia nos conjuntos de treinamento e validação; (b) exibe a progressão da perda durante o treinamento e validação nos mesmos conjuntos.

5.2. Aprendizagem de Conjunto (Comitê)

Apesar dos testes considerarem várias combinações de CNNs, o desempenho do comitê não superou o de uma CNN individual. A partir dos resultados obtidos do comitê, observamos que a média das métricas de avaliação apresentou resultados satisfatórios. A acurácia média foi de 94,35%, indicando que o modelo foi capaz de prever corretamente a classe dos dados em aproximadamente 94,35% das vezes. Além disso, a precisão média foi de 96,89%, o que significa que, em média, cerca de 96,89% das previsões positivas feitas pelo modelo estavam corretas. O *recall* médio foi de 94,35%, indicando a proporção de instâncias positivas que foram corretamente identificadas pelo modelo em relação ao total de instâncias positivas no conjunto de dados. O F1 Score médio, que é uma média harmônica entre precisão e *recall*, foi de 94,76%, demonstrando um bom equilíbrio entre precisão e *recall*.

O coeficiente Kappa médio foi de 0,94, que é uma medida de concordância entre as previsões do modelo e as classes reais, corrigida pelo acaso. O desvio padrão das métricas do comitê é baixo, indicando consistência nos resultados entre diferentes conjuntos de

dados e estabilidade no desempenho do modelo. Individualmente, as melhores redes foram a DenseNet169, Xception e InceptionV3, como mostrado na Tabela 1. O comitê foi formado com base na média ponderada das previsões das CNNs avaliadas neste estudo.

5.3. Comparação com o Estado da Arte

Os resultados fornecem um panorama comparativo entre diversos estudos na área de processamento de imagens de cromossomos, enfatizando a métrica acurácia (A) e apresentando, de forma exclusiva para este trabalho, uma avaliação mais abrangente que inclui o índice de concordância *kappa* (K) e o desvio padrão (DP).

Conforme evidenciado na Seção 2, o estudo de [Xia et al. 2023] alcançou a mais alta acurácia entre todos os estudos listados, atingindo 99,58%. Em seguida, o trabalho de [Lin et al. 2022] obteve uma acurácia de 99,33%, com precisão, *recall* e *F1-Score* todos em 99%. No entanto, a falta de informações sobre o coeficiente de concordância *kappa* e desvio padrão em ambos os estudos deixa uma lacuna na avaliação da consistência e confiabilidade dos modelos. É importante destacar que os estudos de [Xia et al. 2023] e [Lin et al. 2022] utilizaram conjunto de dados com 11.391 e 126.543 imagens de cromossomos, respectivamente.

O presente trabalho, não só alcançou uma acurácia relevante de aproximadamente de 99%, mas também se destacou por utilizar uma abordagem mais holística na avaliação do desempenho do modelo. O trabalho também reporta o índice de concordância *kappa*, atingindo um valor próximo da perfeição (0,99), e desvio padrão baixo (0,002) para todas as métricas, indicando uma alta confiabilidade e consistência nos resultados obtidos.

Em resumo, enquanto os estudos de [Xia et al. 2023] e [Lin et al. 2022] apresentam acurácias elevadas, o presente trabalho se destaca por oferecer uma avaliação compreensiva do modelo, enfatizando não só a acurácia, mas também a precisão, o *recall*, o *F1-Score*, todas aproximando-se de 99%, e, de forma singular, o índice de concordância *kappa* e o desvio padrão. Estes critérios avaliativos adicionais fornecem uma nova perspectiva sobre a confiabilidade e a consistência do modelo, aspectos cruciais para aplicações práticas em cenários do mundo real.

6. Conclusão

Os experimentos de classificação multiclasse no conjunto de dados ChromosomeNet revelaram que a performance dos modelos CNN é significativamente influenciada pela escolha da arquitetura da rede e do otimizador (*ADAM* ou *SGD*) utilizados. A estratégia DFT mostrou-se promissora, independentemente do otimizador utilizado, com todas as arquiteturas alcançando acurácias acima de 90%. Especificamente, com o otimizador *ADAM*, as arquiteturas *DenseNet169*, *Xception*, *Inception*, *DenseNet121* e *EfficientNetV2B3* se destacaram, enquanto com o otimizador *SGD*, a *DenseNet201* exibiu a maior acurácia, precisão, *F1-Score* e coeficiente *Kappa*, além de um baixo desvio padrão, ressaltando a confiabilidade e consistência dos resultados.

A escolha do otimizador teve um impacto significativo na performance dos modelos. Especificamente, o otimizador *ADAM* com a estratégia DFT demonstrou promover maior precisão e estabilidade nos resultados. Isso sugere que a utilização do *ADAM* pode ser uma estratégia eficaz para melhorar a eficácia do modelo em tarefas de classificação complexas.

A pesquisa sobre a classificação de imagens de cromossomos utilizando CNNs revelou limitações que impactam a eficácia e generalização desses modelos. Uma das principais restrições é a sensibilidade à escolha da arquitetura da rede; o impacto significativo que a arquitetura escolhida tem na performance dos modelos é um exemplo disso. Isso demonstra que, para esse tipo de dados, alguns tipos de arquiteturas podem funcionar melhor do que outros, o que limita a generalização. Além disso, a dependência do otimizador e a necessidade de estratégias de treinamento específicas, como a estratégia DFT, dificultam o processo de treinamento e podem limitar o uso geral dos modelos. A generalização dos resultados também é um problema porque exige validação em vários conjuntos de dados ou domínios de imagens de cromossomos, enquanto a complexidade da seleção de componentes enfatiza que uma abordagem cuidadosa e uma compreensão completa do problema são necessárias para obter resultados satisfatórios. Essas limitações mostram que a construção de modelos de classificação de imagens de cromossomos com CNNs enfrenta desafios e indicam que uma investigação abrangente é necessária para superá-los.

Referências

- [Abid and Hamami 2018] Abid, F. and Hamami, L. (2018). A survey of neural network based automated systems for human chromosome classification. *Artificial Intelligence Review*, 49:41–56.
- [Anh et al. 2022] Anh, L., Thanh, V., Son, N., Phuong, D. K., Anh, L., Ram, T., Minh, B., Tung, H., Thinh, N., Ha, V., and Ha, M. (2022). Efficient type and polarity classification of chromosome images using cnns: a primary evaluation on multiple datasets. In *2022 IEEE Ninth International Conference on Communications and Electronics (ICCE)*, pages 400–405.
- [Chollet 2017] Chollet, F. (2017). Xception: Deep learning with depthwise separable convolutions.
- [He et al. 2016] He, K., Zhang, X., Ren, S., and Sun, J. (2016). Deep residual learning for image recognition. In *2016 IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, IEEE Computer Society:770–778.
- [Hsu and Li 2023] Hsu, C.-Y. and Li, W. (2023). Explainable geoi: can saliency maps help interpret artificial intelligence’s learning process? an empirical study on natural feature detection. *37(5):963–987*.
- [Huang et al. 2017] Huang, G., Liu, Z., and Van Der Maaten, L. and Weinberger, K. Q. (2017). Densely connected convolutional networks. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, page 4700–4708.
- [Izadyyazdanabadi et al. 2018] Izadyyazdanabadi, M., Belykh, E., Mooney, M., Martirosyan, N., Eschbacher, J., Nakaji, P., Preul, M. C., and Yang, Y. (2018). Convolutional neural networks: ensemble modeling, fine-tuning and unsupervised semantic localization for neurosurgical cle images. 54.
- [Kornblith et al. 2019] Kornblith, S., Shlens, J., and Le, Q. V. (2019). Do better imagenet models transfer better? *Computer Vision and Pattern Recognition (cs.CV)*, 1:2661–2671.

- [Lin et al. 2022] Lin, C., Chen, H., Huang, J., Peng, J., Guo, L., Yang, Z., Du, J., Li, S., Yin, A., and Zhao, G. (2022). Chromosomenet: A massive dataset enabling benchmarking and building baselines of clinical chromosome classification. *Computational Biology and Chemistry*, 100:107731.
- [Moon et al. 2020] Moon, W. K., Lee, Y.-W., Ke, H.-H., Lee, S. H., Huang, C.-S., and Chang, R.-F. (2020). Computer-aided diagnosis of breast ultrasound images using ensemble learning from convolutional neural networks. *Computer Methods and Programs in Biomedicine*, 190:105361.
- [Perez and Wang 2017] Perez, L. and Wang, J. (2017). The effectiveness of data augmentation in image classification using deep learning. *Computer Vision and Pattern Recognition (cs.CV)*.
- [Russakovsky et al. 2015] Russakovsky, O., Deng, J., Su, H., Krause, J., Satheesh, S., Ma, S., Huang, Z., Karpathy, A., Khosla, A., and Bernstein, M. (2015). Imagenet large scale visual recognition challenge. *International Journal of Computer Vision*, 115(3), page 211–252.
- [Santos et al. 2021] Santos, F., Veras, R., Santos, E., Claro, M., Vogado, L. H., Ito, M., and Bianchi, A. (2021). Uma avaliação de arquiteturas de aprendizado profundo para a classificação de Úlceras do pé diabético. In *Anais do XXI Simpósio Brasileiro de Computação Aplicada à Saúde*, pages 323–334, Porto Alegre, RS, Brasil. SBC.
- [Saranya and Lakshmi 2023] Saranya, S. and Lakshmi, S. (2023). Classification of chromosomes to diagnose chromosomal abnormalities using cnn. In *2023 International Conference on Artificial Intelligence and Knowledge Discovery in Concurrent Engineering (ICECONF)*, pages 1–5.
- [Simonyan and Zisserman 2014] Simonyan, K. and Zisserman, A. (2014). Very deep convolutional networks for large-scale image recognition. *Computer Vision and Pattern Recognition (cs.CV)*, 1.
- [Sokolova and Lapalme 2009] Sokolova, M. and Lapalme, G. (2009). A systematic analysis of performance measures for classification tasks. *Information Processing Management*, 45(4):427–437.
- [Szegedy et al. 2016] Szegedy, C., Vanhoucke, V., Ioffe, S., Shlens, J., and Wojna, Z. (2016). Rethinking the inception architecture for computer vision. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, page 2818–2826.
- [Tan and Le 2021] Tan, M. and Le, Q. V. (2021). Efficientnetv2: Smaller models and faster training. *Computer Vision and Pattern Recognition (cs.CV) (ICML 2021)*, 1.
- [Vogado et al. 2021] Vogado, L., Veras, R., Aires, K., A. F., Silva, R., Ponti, M., and Tavares, J. M. R. (2021). Diagnosis of leukaemia in blood slides based on a fine-tuned and highly generalisable deep learning model. *Sensors*, 21(9), page 2989–2989.
- [Xia et al. 2023] Xia, C., Wang, J., Qin, Y., Wen, J., Liu, Z., Song, N., Wu, L., Chen, B., Gu, Y., and Yang, J. (2023). Karyonet: Chromosome recognition with end-to-end combinatorial optimization network. *IEEE Transactions on Medical Imaging*, 42(10):2899–2911.