

UNIVERSIDADE FEDERAL DE SÃO CARLOS

**TRABALHO DE CONCLUSÃO DE CURSO
ENGENHARIA ELÉTRICA**

**IMPLEMENTAÇÃO E AVALIAÇÃO DE REDES NEURAIS COMPACTAS
PARA DETECÇÃO DE CATARATA COM DADOS LIMITADOS**

Aluno: Felipe Estrada Nunes da Silva

Orientador: Celso Aparecido de França

São Carlos, 28 de Fevereiro de 2025

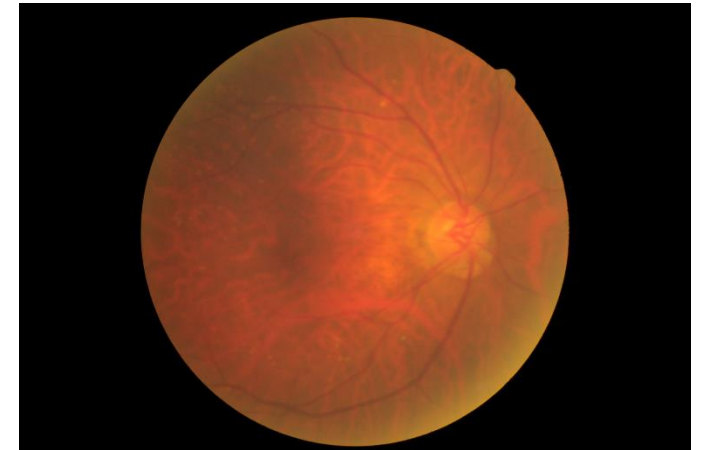
CONTEXTUALIZAÇÃO

Definição da Catarata:

- Opacificação do cristalino, reduzindo a passagem de luz;
- Principal causa de cegueira reversível no mundo;
- Associada ao envelhecimento, diabetes e fatores genéticos;

Impacto da Catarata:

- Redução significativa na qualidade de vida;
- Dificuldade para ler, dirigir e realizar atividades diárias;
- Cirurgia é a única solução definitiva, mas diagnóstico precoce pode retardar sua progressão;



JUSTIFICATIVA DO TRABALHO

- Métodos tradicionais de diagnóstico dependem de exames oftalmológicos presenciais;
- Demanda por soluções automatizadas para detecção precoce;
- Inteligência Artificial pode acelerar diagnósticos e reduzir custos médicos;

OBJETIVO

- Implementar e avaliar diferentes redes neurais compactas para detecção de catarata;
- Trabalhar com bases de dados limitadas para simular cenários clínicos reais;
- Foco em situações com baixa capacidade de processamento;
- Comparar arquiteturas e desempenhos de CNN, Vision Transformers e modelos híbridos;

INTELIGÊNCIA ARTIFICIAL

Machine Learning (ML)

- Ramo da inteligência artificial que permite que algoritmos aprendam padrões a partir de dados, sem a necessidade de programação explícita.

Deep Learning (DL)

- Subcampo do Machine Learning baseado em redes neurais profundas. Essas redes possuem múltiplas camadas que aprendem representações complexas dos dados.

Redes Neurais Convolucionais (CNNs)

- Foco em dados bidimensionais;
- Excelentes para análise de imagens;
- Extraem características como bordas, texturas e padrões visuais;
- Utilizadas amplamente na área médica para segmentação e diagnóstico;

REDES NEURAIS CONVOLUCIONAIS (CNN)

Camada de Convolução

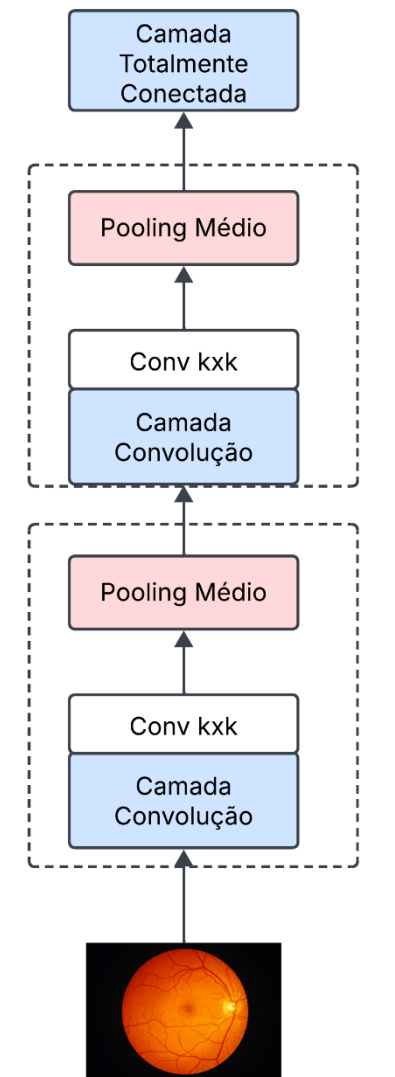
- Matrizes numéricas conhecidas como filtros, ou Kernels;
- Soma do produto dos elementos do filtro com a entrada;
- Características dependentes do tipo do filtro;
- Aplicação função de ativação, ReLU por exemplo;

Camada de Subamostragem (Pooling)

- Reduzir dimensão de dados;
- Diversos tipos de pooling (max, average);

Camada Totalmente Conectada

- Última camada;
- Etapa de conexão entre neurônios e classificação



EFFICIENTNET

Rede Neural Convolucional

Introdução de aumento de parâmetros proporcionalmente

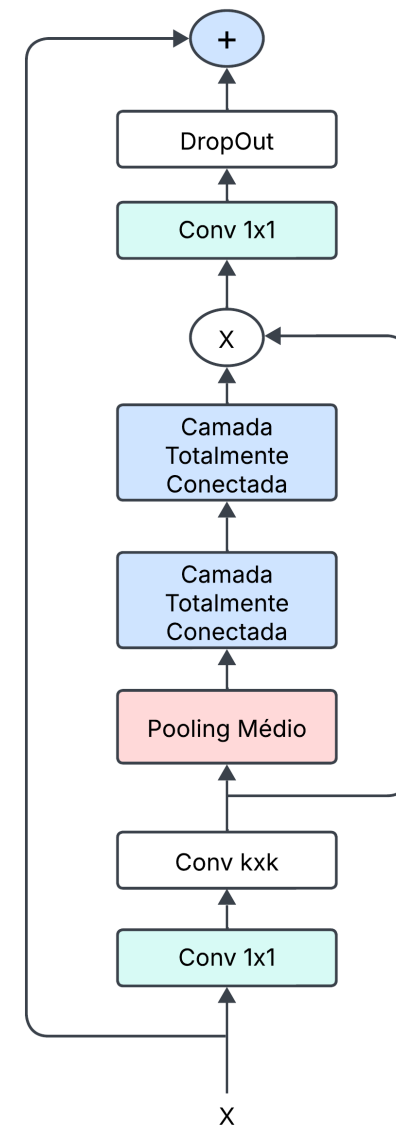
- Profundidade (nº de camadas);
- Largura (nº canais);
- Resolução;

Mobile Inverted Bottleneck (MBConv)

- Expande canais antes da convolução depthwise;
- Reduz custo computacional e melhora desempenho em redes neurais móveis;

Squeeze-and-Excitation

- Recalibração de ajustes dos pesos de diferentes canais;



RESNET

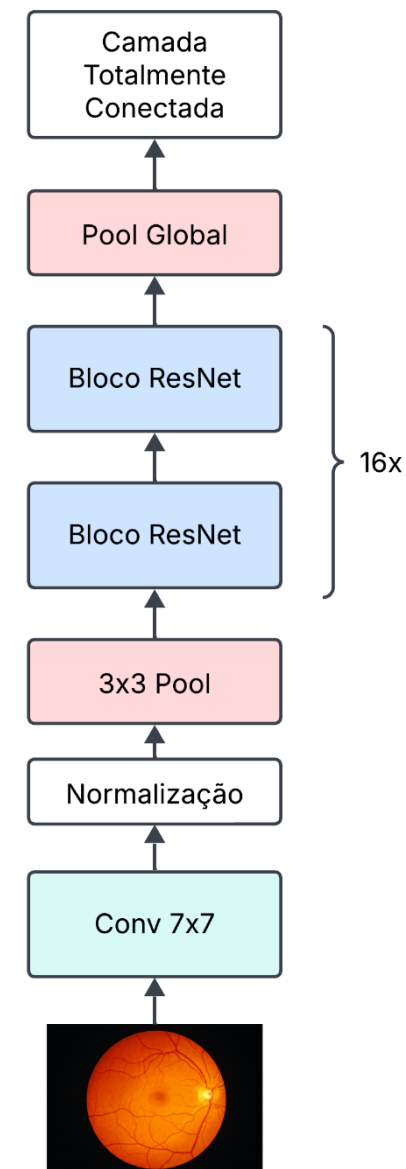
Rede Neural Convolucional

Introdução de conexões residuais

- Soma da entrada a saída das camadas de convolução;

Blocos ResNet

- Convolução;
- Normalização;
- ReLU;
-
- ResNet50: 3 convoluções por bloco;
- ResNet18: 2 convoluções por bloco;
- ResNet8: 1 convoluções por bloco;



TRANSFORMER

- Criada para tarefas de processamento de linguagem natural (NLP);
- Introduziu o mecanismo de atenção e atenção própria;
- Eficiência em tarefas de processamento sequencial;
- Capacidade de paralelização;

Camada de Embedding

- Vetorização da entrada;

Camada Posicional

- Inserção de valor posicional ao vetor;

Mecanismo de Atenção (Self-Attention)

- Cálculo da relevância de entre partes da entrada;
- Utiliza projeções lineares Q (consulta), K (chaves) e V (valores);
- Cálculo via multiplicação com pesos aprendíveis;
- Produto escalar entre Q e K determina similaridade entre elementos;

TRANSFORMER

Atenção Multicabeça

- Múltiplos mecanismos de atenção;

Feedforward

- Camadas totalmente conectadas que armazenam conhecimento;

Normalização de Camada

- Normalização da saída para redução de instabilidade;

Camada de Saída

- Integração de informações processadas para tarefa final;

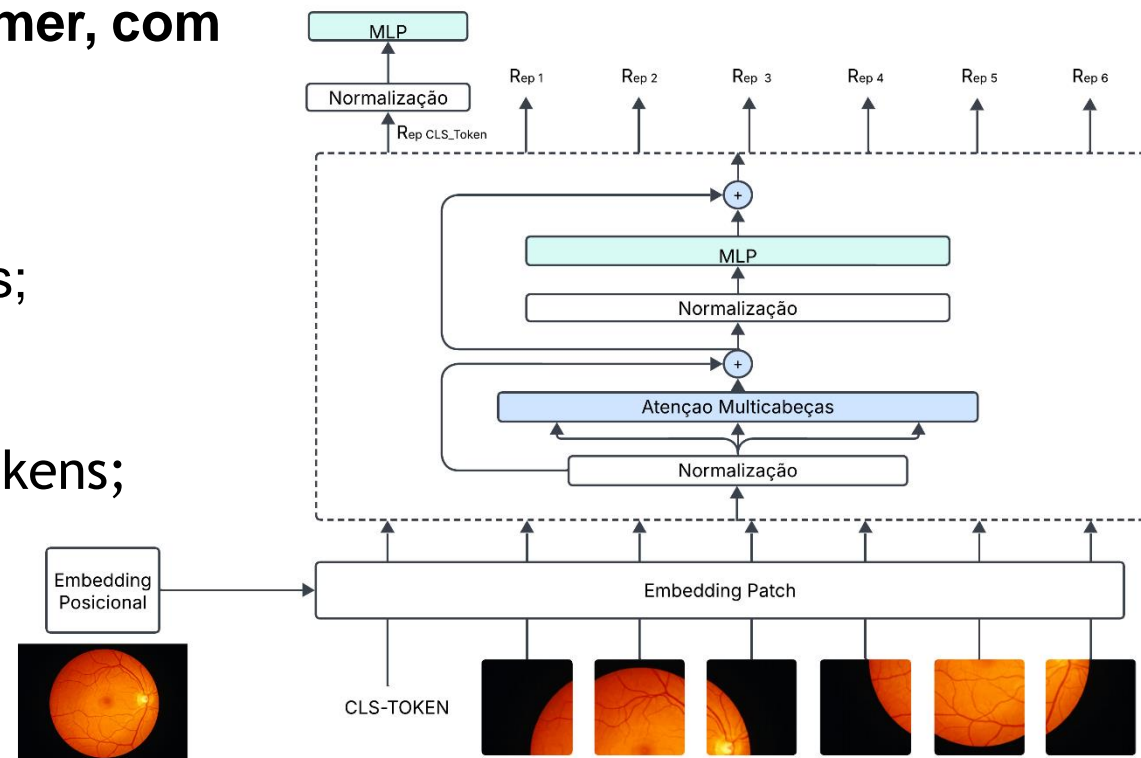
VISION TRANSFORMER (ViT)

Arquitetura Transformer

- Processamento de imagens em batches menores;
- Utilizado com frequência em grandes bases de dados;

Segue mesmas camadas do Transformer, com exceção:

- Patch Embedding
Vetorização da entrada em tokens;
Batches com tamanhos fixos;
- Token de Classe
Agrega informações dos outros tokens;



METODOLOGIA

Dados utilizados

- Dataset especializado em detecção de doenças oculares;
- Categorias: Normal, catarata, glaucoma, doenças de retina;
- Binária: Normal e catarata;
- Multiclasses: 4 classes;

Balanceamento e Aumento de Dados

- Data augmentation;

Pré-processamento de Imagens

- Redimensionamento de imagens;
- Normalização;

Particionamento do Dataset

- Treinamento (80%);
- Validação (12%);
- Teste (8%);

MODELAGEM

Adaptação de ultimas camadas para quantidade de classes

EfficientNet-B0

- Versão compacta da EfficientNet;

ResNet

- ResNet6 e ResNet8: Uma camada convolucional por bloco;
- ResNet10, e ResNet18: Duas camadas convolucionais por bloco;

Vision Transformer Lite (ViT-Lite)

- Redução de camadas no encoder (12 para 4);
- Redução de cabeças de atenção (12 para 2);
- Implementação do pooling sequencial;

MODELAGEM

Compact Convolutional Transformer (CCT)

- ViT-Lite com convolução na entrada (Patch Embedding);

ResNet + ViT

- ResNet como tokenizador no lugar do Patch Embedding;
- ResNet para extração de características;
- Encoder ViT
- Desativação aleatória de neurônios durante o treinamento (DropOut) mais robusto (30%)
- Pruning para reduzir dimensionalidade;

TREINAMENTO

Python + Pytorch

Hiper parâmetros

- AdamW;
- taxa de aprendizado inicial $4 \cdot e^{-6}$;
- Redução de taxa de aprendizado (ReduceLROnPlateau);
- Backpropagation;
- Early Stopping;

Métricas

- Acurácia;
- Perda CrossEntropyLoss;
- Sensibilidade;

$$\text{Acurácia} = \frac{N_{\text{Previsões Corretas}}}{\text{Total de amostras}} \cdot 100 \%$$

$$L = \frac{-1}{N} \cdot \sum_{i=1}^N \sum_{c=1}^C y_{i,c} \log p_{i,c}$$

$$\text{Sensibilidade} = \frac{VP}{VP + FN} \cdot 100 \%$$

DESEMPENHO DOS MODELOS

	Numero Parâmetros	Total Modelo (MB)	Precisão (%)		Perda		Sensibilidade		Tempo de Inferência (ms)	
			Binário	Multiclasse	Binário	Multiclasse	Binário	Multiclasse	Binário	Multiclasse
EfficientNet-B0	4,013 M	47,49	95,31	59,38	0,2641	0,9888	95,45%	86,27%	610,783	347,99
ResNet-18	11,178 M	54,28	100	97,66	0,0308	0,1408	100,00%	98,28%	226,64	341,096
ResNet-10	4,908 M	27,77	100	93,75	0,0308	0,225	100,00%	98,00%	163,379	406,335
ResNet-8	4,734 M	26,04	100	96,09	0,0084	0,1033	100,00%	98,73%	144,344	102,381
ResNet 6	1,599 M	11,97	100	89,84	0,0583	0,3989	100,00%	96,73%	97,589	179,487
CVT	0,396 M	3912,58	96,88	66,41	0,1037	0,7928	98,08%	88,85%	4028,623	2047,364
CCT	0,631 M	30,76	98,44	64,84	0,1046	0,7968	98,08%	88,26%	3401,841	2053,903
ResNet18-ViT	12,036 M	56,69	100	98,44	0,0308	0,0686	100,00%	99,49%	166,331	384,676
ResNet10-ViT	5,765 M	29,05	100	95,31	0,0308	0,1513	100,00%	98,51%	180,752	321,327
ResNet8-ViT	5,117 M	27,15	100	97,66	0,0018	0,0938	100,00%	99,21%	146,805	212,937
ResNet06-ViT	1,981 M	13,08	100	95,31	0,0015	0,3407	100,00%	98,45%	133,271	63,442

CLASSIFICAÇÃO BINÁRIA

ResNet e ResNet-ViT

- 100% de precisão no teste;
- Precisão acima de 95% em validação;
- Rápida conversão no treinamento (10 épocas);

CCT

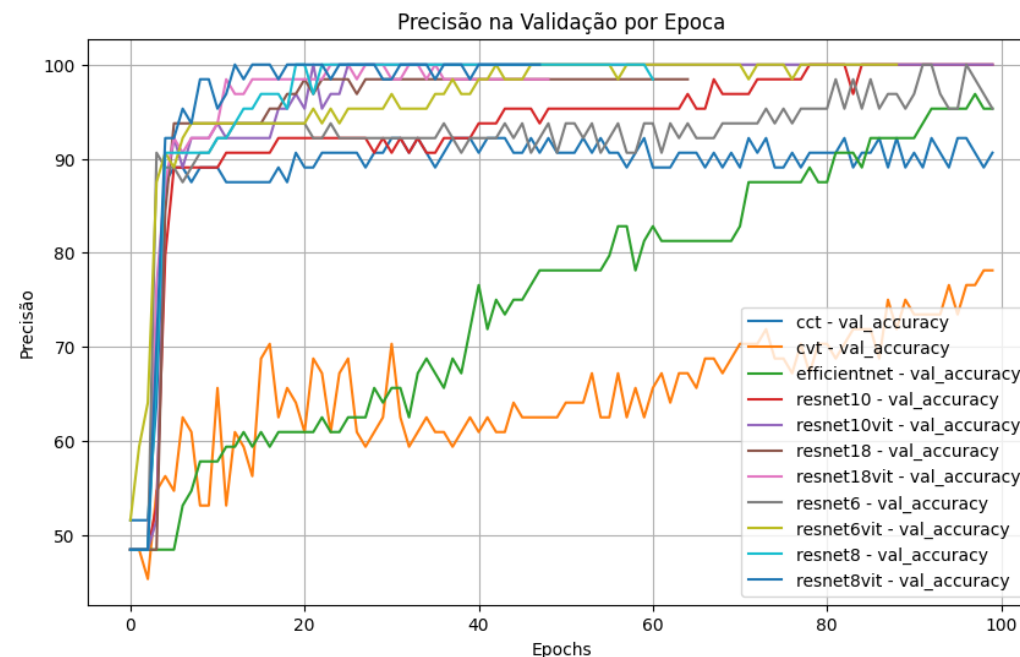
- 98,44% de precisão no teste;
- Precisão acima de 90% em validação;
- Rápida conversão no treinamento (20 épocas);

CVT

- 96,88% de precisão no teste;
- Precisão baixa (78%) em validação;

EfficientNet

- 95,31% de precisão no teste;
- Precisão acima de 95% em validação;
- Conversão lenta no treinamento (100 épocas);



CLASSIFICAÇÃO MULTICLASSES

ResNetViT

- Acima de 95,31% de precisão no teste;
- Precisão acima de 90% em validação;
- Rápida conversão no treinamento (20~60 épocas);

ResNet

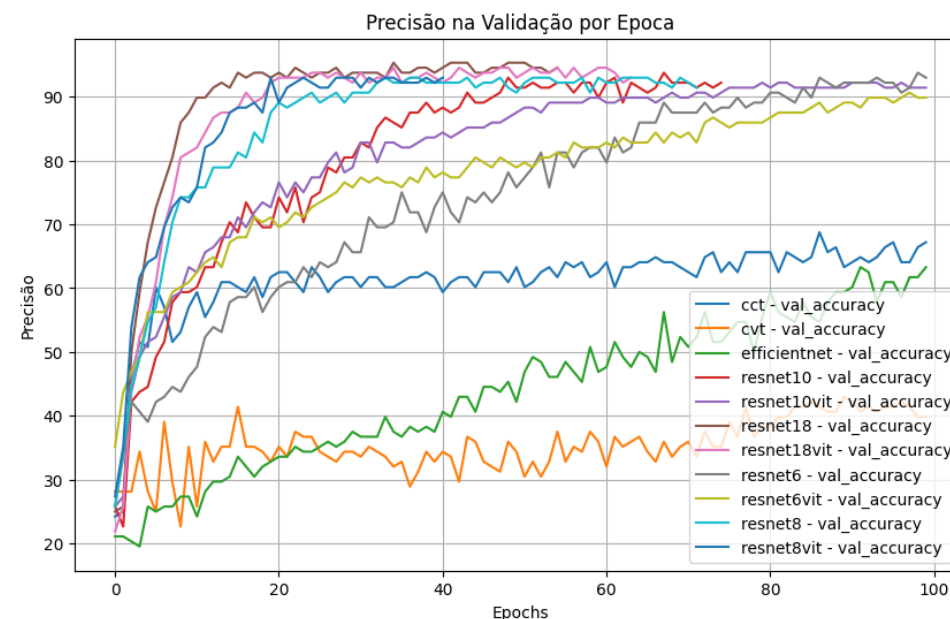
- Acima de 89,84% de precisão no teste;
- Precisão acima de 90% em validação;
- Rápida conversão no treinamento (40~80 épocas);

CVT e CCT

- 64,84% e 66,41% de precisão no teste;
- Precisão abaixo de 70% em validação;

Efficient

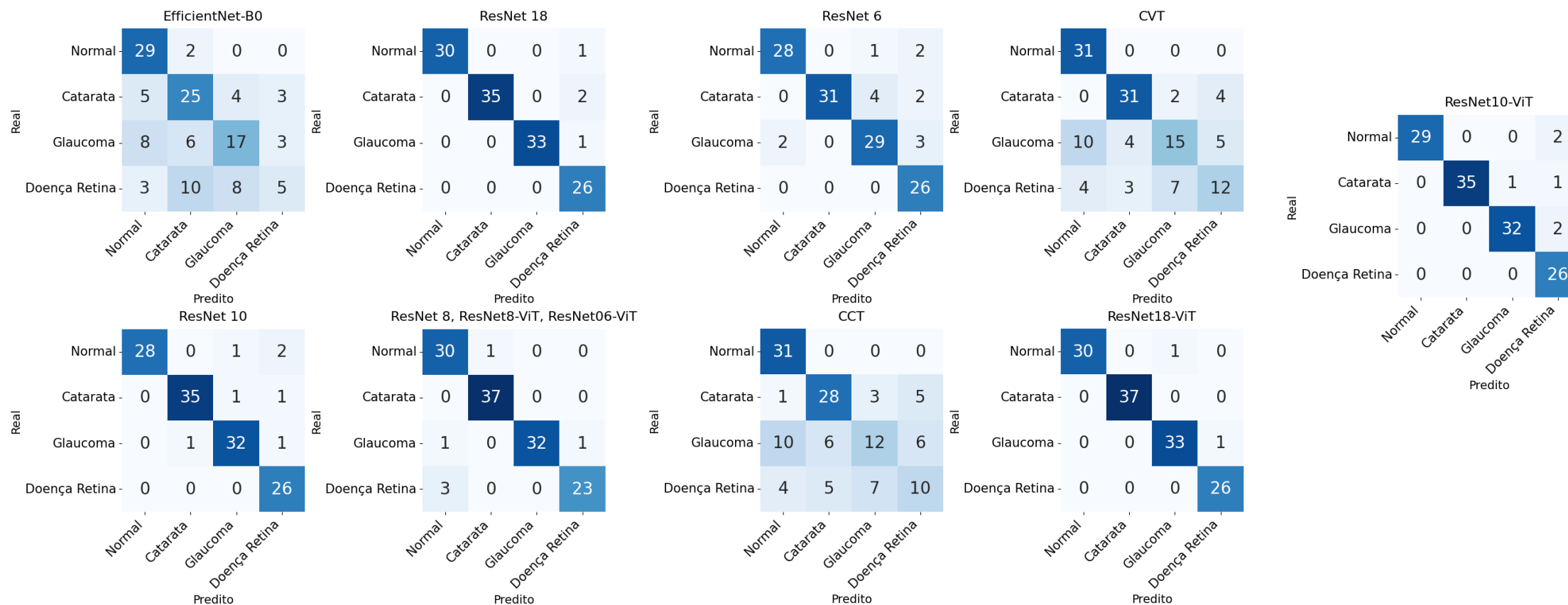
- 59,38% de precisão no teste;
- Precisão baixa (~60%) em validação;
- Conversão de aprendizado lenta;



CLASSIFICAÇÃO MULTICLASSES

Análise de Resultados

- Baixo impacto dos ViT em modelos híbridos;



CONCLUSÕES

- Redes neurais compactas são viáveis e eficazes;
- Como esperado, baixo impacto dos Transformers em modelos híbridos;
- Necessário avaliação de aplicação para justificar utilização de modelos híbridos;
- Modelos baseados apenas em Transformers tiveram desempenho bem abaixo, por conta da limitação do tamanho do banco de dados;