Destilação de Conhecimento Evidencial Adaptativa Empregada na Classificação de Imagens da Cavidade Oral

Carlos Alberto Matias de Abreu Júnior[1], **Thiago Pirola Ribeiro**[1], **Marcelo Zanchetta do Nascimento**[1]

[1]Faculdade de Computação – Universidade Federal de Uberlândia (UFU)
Uberlândia - MG – Brazil

carlos.alberto@ufu.br, tpribeiro@ufu.br, marcelo.nascimento@ufu.br

Abstract. *Oral cancer presents survival rates above 80% when diagnosed at early stages; however, this rate drops to below 20% in advanced disease. Despite recent advances demonstrating high performance in digital slide classification, the clinical adoption of these approaches remains limited by relevant methodological challenges. This work proposes an A-EKD methodology for the reliable diagnosis of OSCC via evidential knowledge distillation. The Gate ($\mathcal{G}$) is introduced to filter Teacher uncertainties, protecting the Student MobileNetV3's learning in noisy samples. Under experimental rigor to avoid data leakage, the model achieved 90.98% accuracy, surpassing the state-of-the-art of 87.1%. Statistical analysis shows a classification potential supported by an Odds Ratio of 24.5 and an Err-AUC of 0.8490, validating the premise of filtering uncertainties. The method mitigates overconfidence and optimizes calibration, enabling reliable diagnoses through lightweight networks, which possess a smaller number of parameters.*

Resumo. *O câncer oral apresenta taxas de sobrevivência superiores a 80% quando diagnosticado precocemente; contudo, esse índice cai para menos de 20% em estágios avançados. Apesar dos avanços recentes indicarem elevado desempenho na classificação de lâminas digitais, a adoção clínica dessas abordagens ainda é limitada por desafios metodológicos relevantes. Este trabalho propõe uma metodologia A-EKD para o diagnóstico confiável de OSCC via destilação de conhecimento evidencial. Introduz-se o Gate ($\mathcal{G}$) para filtrar incertezas do Professor, protegendo o aprendizado do Aluno MobileNetV3 em amostras ruidosas. Sob rigor experimental para evitar vazamento de dados, o modelo atingiu 90,98% de acurácia, superando o estado da arte de 87,1%. A análise estatística mostra um potencial da classificação embasado por um Odds Ratio de 24,5 e uma Err-AUC de 0,8490, validando a premissa de filtrar incertezas. O método mitiga a superconfiança e otimiza a calibração, viabilizando diagnósticos confiáveis através de redes leves, que possuem um menor número de parâmetros.*

1. Introdução

O câncer oral apresenta taxa de sobrevivência superior a 80% quando diagnosticado precocemente. Entretanto, esse prognóstico declina de forma acentuada para menos de 20% em estágios avançados. O Carcinoma de Células Escamosas Oral (do inglês, *Oral*

Squamous Cell Carcinoma – OSCC), responsável por mais de 85% dos casos, exemplifica a gravidade desse cenário, uma vez que a maioria dos pacientes ainda é diagnosticada tardiamente. Além disso, a aparência clínica das lesões é frequentemente insuficiente para diferenciar o grau de displasia ou a invasão tumoral, dificultando a definição imediata da conduta terapêutica e tornando a confirmação histopatológica indispensável [WHO Classification of Tumours Editorial Board 2024].

Embora avanços recentes em inteligência artificial tenham demonstrado alta acurácia na classificação de lâminas digitais, a implementação clínica dessas ferramentas enfrenta desafios críticos relacionados ao *data leakage* e à falta de validade externa em conjuntos de dados independentes [Nogueira and Gomes 2025, Paraíso and Machado 2025]. Nesse contexto, a necessidade de sistemas automatizados que não apenas classifiquem, mas também forneçam métricas de confiabilidade sobre suas predições, tornou-se um requisito ético e técnico para a patologia digital moderna [Nogueira and Gomes 2025, Agresti 2012].

A modelagem da ambiguidade através da Aprendizagem Evidencial Profunda (EDL) surgiu como uma alternativa robusta ao uso de funções *softmax* convencionais, permitindo a quantificação da incerteza evidencial (*vacuity*) como uma medida de incerteza epistêmica [Xiang et al. 2025, Guo et al. 2017]. No entanto, abordagens recentes de destilação de conhecimento, como o método EKD proposto por [Xiang et al. 2025], embora inovadoras ao transferir distribuições de segunda ordem, ainda apresentam riscos de corrupção do aprendizado do aluno quando o modelo professor manifesta incertezas em amostras ruidosas. Para mitigar essa propagação de ruído, a presente investigação introduz um mecanismo de filtragem adaptativa através da métrica de confiabilidade ($\mathcal{G}$), que monitora a incerteza do professor para assegurar que apenas conhecimentos de alta confiança sejam destilados para a rede menor [Xiang et al. 2025, Paraíso and Machado 2025]. Adicionalmente, o desenvolvimento do método A-EKD (Destilação de Conhecimento Evidencial Adaptativa) aborda diretamente o problema da superconfiança (*overconfidence*), um fenômeno documentado em arquiteturas profundas que tendem a atribuir altas probabilidades a diagnósticos incorretos [Guo et al. 2017, Paraíso and Machado 2025]. Ao otimizar o Erro de Calibração Esperado (ECE), a metodologia proposta busca alinhar a confiança teórica do modelo com sua acurácia empírica, permitindo a criação de modelos mais compactos e estatisticamente "honestos" sem sacrificar a sensibilidade diagnóstica necessária para o câncer oral [Guo et al. 2017, Agresti 2012].

Finalmente, a viabilização dessas técnicas em arquiteturas eficientes como a *MobileNetV3* responde a uma demanda global por soluções de baixo custo que possam ser integradas a dispositivos móveis e sistemas de telemedicina em tempo real [Paraíso and Machado 2025, Nogueira and Gomes 2025]. Ao reduzir a barreira tecnológica e financeira para o processamento de imagens histopatológicas, o A-EKD visa democratizar o acesso ao diagnóstico precoce do OSCC em regiões de alta vulnerabilidade socioeconômica, transformando o rigor da análise estatística em um impacto social direto na redução da morbidade oncológica [Nogueira and Gomes 2025, Paraíso and Machado 2025].

O presente trabalho propõe avançar na classificação convencional, aprimorando a integração de EL e a KD proposta por [Xiang et al. 2025]. Diferente de técnicas

que buscam apenas padrões visuais, a abordagem proposta utiliza a modelagem de incerteza para captar informações sobre a confiabilidade dos dados e resistir a ruídos e variações de escala (zoom de 100 ou 400 vezes). Sistemas computacionais capacitados com essa arquitetura híbrida podem ser explorados para processar múltiplos níveis de abstração de forma eficiente, o que pode resultar em uma interpretação de imagens que não apenas reduz o tempo e esforço, mas identifica os elementos essenciais para um diagnóstico médico [Prajwal et al. 2025, Lambert et al. 2024, Zhang et al. 2025]. As principais contribuições deste trabalho são:

- Investigação do *Gate* ($\mathcal{G}$), que utiliza a vacuidade evidencial do Professor para filtrar a transferência de conhecimento.
- Estudo de um método A-EKD (Destilação de Conhecimento Evidencial Adaptativa) que explora obter o menor Erro de Calibração Esperado (ECE), o que pode analisar modelos mais compacto;
- Análise da viabilização de diagnósticos em arquiteturas como a *MobileNetV3* para sistemas de baixo custo.

2. Materiais e Métodos

A metodologia proposta fundamenta-se na junção das técnicas EL e KD para análise em modelos como EfficienteNetB0 e MobileNetV3Small. A etapa de pré-processamento segue os mesmos passos propostos por [Nogueira and Gomes 2025], que foram aplicados a base de imagens também investigada neste trabalho. Esta seção detalha o processo executado para treinamento e classificação dos dados, assim como, as arquiteturas selecionadas e a definição do algoritmo de EKD proposto por [Xiang et al. 2025]. Na Figura 1 é apresentado o fluxograma da metodologia proposta. Neste estudo empregou-se o paradigma *Teacher-Student*, no qual uma rede robusta transfere conhecimento para uma rede compacta. Os valores adotados nos modelos são descritos na Tabela 1.

2.1. Conjunto de Dados e Pré-processamento

As amostras histológicas foram preparadas com coloração H&E e catalogadas por especialistas, sendo posteriormente digitalizadas via microscópio Leica ICC50 HD. O banco de dados resultante totaliza 1.224 micrografias provenientes de 230 indivíduos, divididas em grupos de magnificação de $100x$ (439 OSCC e 89 normais) e $400x$ (495 OSCC e 201 normais). Globalmente, o conjunto de dados apresenta 934 imagens classificadas como malignas (OSCC) e 290 referentes ao epitélio normal da cavidade oral [Rahman et al. 2020]. Na Figura 2 são apresentados exemplos de tecidos normais e malignos (OSCC).

Para tratar o viés de classe e garantir a generalização do modelo, aplicou-se um protocolo seguido por [Nogueira and Gomes 2025]: i) Particionamento Estratificado: O dataset original foi dividido em subconjuntos de Validação (10%) e Teste (20%) e o restante das amostras foi Treino. A proporção de classes foi preservada, e as porcentagens foram definidas considerando a classe com menor número de amostras; ii) Balanceamento via *Data Augmentation*: Uma vez que o particionamento dos dados considerou a classe com menor volume, identificou-se um desequilíbrio significativo para a etapa de treino. Diante disso, na classe definida como "Normal" (minoritária) foram aplicadas

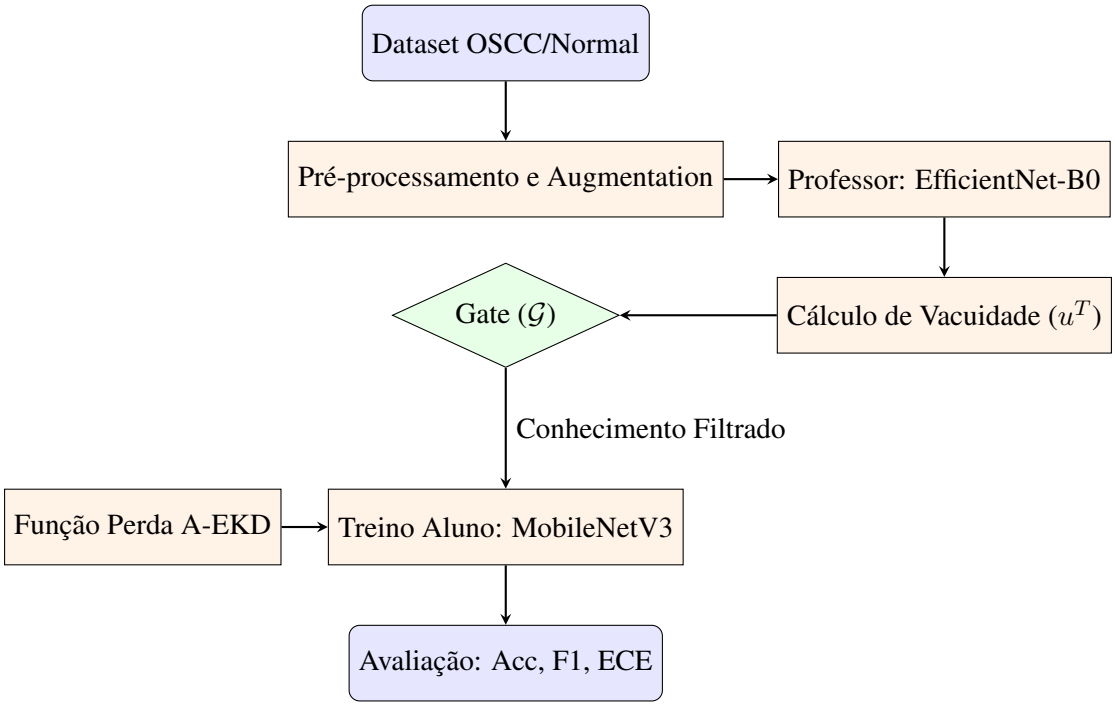
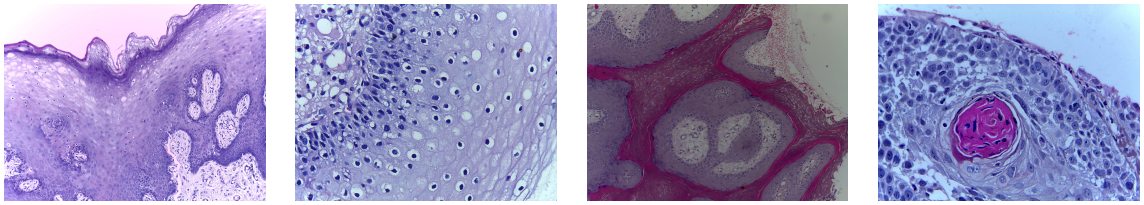


Figura 1. Fluxograma metodológico do A-EKD: Processamento de imagens histopatológicas e destilação adaptativa baseada em incerteza.



(a) Normal 100x **(b) Normal 400x** **(c) OSCC 100x** **(d) OSCC 400x**

Figura 2. Comparativo compacto de lâminas histopatológicas (Normal vs OSCC).

transformações sintéticas geradas via *data augmentation* (rotação aleatória $\pm 20°$, espelhamento horizontal/vertical e ajustes de brilho) exclusivamente às amostras da classe minoritária até que a paridade numérica com a classe "OSCC" fosse atingida. Os conjuntos de validação e teste permaneceram inalterados para garantir a fidelidade da avaliação [Paraíso and Machado 2025]; iii) Normalização: Todas as imagens foram redimensionadas para 224×224 pixels e normalizadas utilizando a média e desvio padrão do ImageNet ($\mu = [0.485, 0.456, 0.406]$, $\sigma = [0.229, 0.224, 0.225]$), facilitando a convergência do treinamento por transferência.

2.2. O Modelo EfficientNet-B0

Para o modelo Professor, selecionou-se a EfficientNet-B0 [Tan and Le 2019], em que esta escolha justifica-se pelo uso do método de *Compound Scaling*, que otimiza uniformemente a profundidade, largura e resolução do modelo. Com aproximadamente 5,3 milhões de parâmetros, a EfficientNet oferece uma representação de características rica e robusta, servindo como uma fonte confiável de "Conhecimento Escuro" (*Dark Knowledge*). Os coeficientes de destilação e fator de sensibilidade utilizados estão na Tabela

1.

2.3. O Modelo MobileNetV3-Small

Para o modelo Aluno, alvo da implantação em dispositivos com baixa capacidade de processamento, utilizou-se a MobileNetV3-Small [Howard et al. 2019]. Projetada via *Neural Architecture Search* (NAS), esta rede emprega blocos de convolução separáveis em profundidade e funções de ativação otimizadas (*hard-swish*), reduzindo o custo computacional para cerca de 2,5 milhões de parâmetros.

Tabela 1. Especificações Técnicas e Hiperparâmetros de Treinamento

Parâmetro	**Configuração / Valor Real**
Arquitetura Professor	EfficientNet-B0 (Pré-treinada)
Arquitetura Aluno	MobileNetV3 Small (Pré-treinada)
Otimizador	Adam
Taxa de Aprendizado (*Learning Rate*)	1×10^{-4}
Função de Perda	A-EKD (Proposta)
Tamanho do Lote (*Batch Size*)	32
Número de Épocas	50
Resolução de Entrada	$224 \times 224 \times 3$
Fator de Sensibilidade da métrica adaptativa (λ)	2.0
Coeficientes de Destilação (β, γ)	1.0, 0.1

Para adaptar as arquiteturas originais ao problema de classificação de OSCC, as camadas de classificação (*heads*) foram reestruturadas. Na Tabela 2 está detalhada a organização das camadas finais do modelo Aluno utilizado no experimento, seguindo a lógica de camadas densas e regularização por *Dropout*.

Tabela 2. Organização de Camadas do Modelo Aluno (MobileNetV3)

Camada (Layer)	**Tipo (Type)**
Entrada (*Input*)	Imagem RGB (224×224)
Backbone	MobileNetV3 Feature Extractor
Global Pooling	Average Pooling
Dense (Linear)	1024 neurônios, ReLU
Dropout	Taxa de 0.2
Dense (Output)	2 neurônios (Evidencial)

2.4. Fundamentação Evidencial de Segunda Ordem

Diferente das redes neurais clássicas que utilizam a função Softmax para colapsar informações em uma estimativa pontual, neste trabalho foi adotada a Aprendizagem Evidencial Profunda (EDL) proposta por [Sensoy et al. 2018]. Nesta abordagem, a saída da rede neural parametriza uma densidade de probabilidade sobre o simplex categórico. Seja $\mathbf{e} = f(\mathbf{x}; \theta)$ o vetor de evidências não-negativas para K classes. Os parâmetros da distribuição de Dirichlet correspondente são definidos por:

$$\boldsymbol{\alpha} = \mathbf{e} + \mathbf{1}, \tag{1}$$

onde a densidade de probabilidade para um vetor de probabilidade $\mathbf{p}$ é dada pela função de Dirichlet [Sensoy et al. 2018]:

$$Dir(\mathbf{p}|\boldsymbol{\alpha}) = \frac{1}{B(\boldsymbol{\alpha})} \prod_{k=1}^{K} p_k^{\alpha_k - 1}, \tag{2}$$

A força total da distribuição, $S = \sum_{k=1}^{K} \alpha_k$, permite a quantificação direta da incerteza epistêmica (ou vacuidade) u através da relação:

$$u = \frac{K}{S}. \tag{3}$$

Esta métrica é fundamental para o mecanismo adaptativo proposto, pois reflete a ausência de evidência teórica do modelo frente a uma amostra específica.

2.5. Metrica adaptativa proposta ($\mathcal{G}$)

O estudo traz a métrica adaptativa $\mathcal{G}(\mathbf{x})$ para mitigar a transferência negativa proveniente de um Professor (*EfficientNet-B0* [Tan and Le 2019]) potencialmente incerto. A métrica modula a intensidade do gradiente de destilação para o Aluno (*MobileNetV3* [Howard et al. 2019]) com base na incerteza do Professor u^T:

$$\mathcal{G}(\mathbf{x}) = \left(1 - u^T(\mathbf{x})\right)^{\lambda} \tag{4}$$

onde λ é um fator de sensibilidade. Esta modulação assegura que o conhecimento do Professor seja transferido apenas quando este demonstrar alta confiança evidencial.

2.6. Otimização Híbrida

A função de perda total para o treinamento do Aluno é definida pela integração da supervisão direta do *ground truth* e da destilação ponderada:

$$\mathcal{L}_{total} = \mathcal{L}_{EDL}(\mathbf{y}, \boldsymbol{\alpha}^S) + \beta \cdot \mathcal{G}(\mathbf{x}) \cdot [\mathcal{L}_{1st} + \gamma \mathcal{L}_{2nd}] \tag{5}$$

O termo $\mathcal{L}_{EDL}$ garante a aderência aos rótulos reais seguindo a formulação de Sensoy et al. [Sensoy et al. 2018], enquanto os termos $\mathcal{L}_{1st}$ e $\mathcal{L}_{2nd}$ promovem o refinamento da calibração estrutural baseada no Professor [Xiang et al. 2025], filtrada pela confiabilidade adaptativa $\mathcal{G}$.

3. Métricas de Avaliação

A validação da metodologia A-EKD fundamenta-se em desempenho preditivo, calibração probabilística, detecção de falhas e significância estatística. Esta abordagem multidimensional explora características do modelo MobileNetV3 destilado em relação a precisão e confiabilidade [Nogueira and Gomes 2025, Paraíso and Machado 2025].

3.1. Desempenho Preditivo

A métrica acurácia (ACC) mede a proporção global de acertos (verdadeiros positivos e negativos) em relação ao total de amostras n:

$$ACC = \frac{TP + TN}{TP + TN + FP + FN}, \tag{6}$$

onde *TP, TN, FP, FN* representam Verdadeiros Positivos, Verdadeiros Negativos, Falsos Positivos e Falsos Negativos, respectivamente.

A medida F1-Score, proposta por [Van Rijsbergen 1975], fornece a média harmônica entre precisão e sensibilidade, sendo importante em cenários de desbalanceamento de classes para evitar a negligência de casos de carcinoma:

$$F1 = \frac{2 \cdot TP}{2 \cdot TP + FP + FN.} \tag{7}$$

3.2. Calibração e Confiabilidade Probabilística

A métrica *Expected Calibration Error* (ECE), introduzida por [Guo et al. 2017], quantifica o desalinhamento entre a confiança prevista e a acurácia observada em M intervalos (*bins*):

$$ECE = \sum_{m=1}^{M} \frac{|B_m|}{n} |acc(B_m) - conf(B_m)|, \tag{8}$$

onde $|B_m|$ é o número de amostras no intervalo m, $acc(B_m)$ é a acurácia real e $conf(B_m)$ é a confiança média do modelo naquele intervalo.

A medida *Brier Score (BS)*, formulada por [Brier 1950], mede o erro quadrático médio das probabilidades previstas f_i em relação aos rótulos binários reais $o_i \in \{0, 1\}$:

$$BS = \frac{1}{n} \sum_{i=1}^{n} (f_i - o_i)^2. \tag{9}$$

Também há a medida *Negative Log-Likelihood (NLL)* que avalia a qualidade do ajuste probabilístico, penalizando predições que apresentam alta confiança em classes incorretas:

$$NLL = -\sum_{i=1}^{n} \sum_{k=1}^{K} y_{i,k} \ln(p_{i,k}), \tag{10}$$

onde $y_{i,k}$ é o indicador binário da classe e $p_{i,k}$ é a probabilidade prevista.

3.3. Análise de Erro e Incerteza

Baseado no protocolo de [Hendrycks and Gimpel 2017], o Erro AUC (Err-AUC) utiliza a vacuidade evidencial (u) como um detector de amostras mal classificadas. O objetivo é que o modelo apresente maiores níveis de incerteza em amostras onde a predição está incorreta, servindo como um mecanismo de segurança para o patologista.

3.4. Significância Estatística

Proposto por [McNemar 1947], esse método é um teste não-paramétrico aplicado as tabelas de contingência 2×2 para comparar modelos correlacionados (Aluno vs. Professor):

$$\chi^2 = \frac{(|b - c| - 1)^2}{b + c}. \tag{11}$$

Para validar a eficácia do modelo A-EKD em comparação ao modelo professor de referência, aplicou-se o Teste de McNemar sobre os pares discordantes de predição. As hipóteses foram:

- Hipótese Nula (H_0): As proporções de discordância entre o modelo Aluno (A-EKD) e o modelo Professor (EfficientNet-B0) são equivalentes ($P_b = P_c$), implicando que não há superioridade estatística entre as arquiteturas no diagnóstico de OSCC.
- Hipótese Alternativa (H_1): As proporções de discordância são significativamente distintas ($P_b \neq P_c$), indicando que o processo de destilação de conhecimento adaptativa resultou em uma mudança sistemática no desempenho preditivo.

O critério de rejeição para H_0 foi estabelecido com um nível de significância de $\alpha = 0,05$.

Também foi calculado o intervalo de confiança (IC 95%) para para delimitar a margem de erro da acurácia p sobre o tamanho da amostra n:

$$IC = 1,96 \cdot \sqrt{\frac{p(1-p)}{n}}. \tag{12}$$

Para quantificar o desempenho do modelo A-EKD em relação ao modelo professor, utiliza-se a Razão de Chances (*Odds Ratio* - OR) aplicada a dados pareados. Diferente do cálculo para grupos independentes, no contexto do teste de McNemar, a métrica foca exclusivamente nos pares discordantes da matriz de confusão. Matematicamente, a fórmula é expressa como:

$$OR = \frac{b}{c}, \tag{13}$$

onde b representa o número de amostras classificadas corretamente apenas pelo modelo proposto (Aluno) e c representa os casos onde apenas o modelo de referência (Professor) obteve o diagnóstico correto. Conforme preconizado por [Agresti 2012], esta métrica fornece uma estimativa direta da força de associação e do ganho de desempenho, indicando quantas vezes é mais provável que o A-EKD acerte um diagnóstico de OSCC nos casos em que as arquiteturas divergem.

4. Resultados e Discussões

Os experimentos foram conduzidos em duas etapas: a primeira em um cenário misto, caracterizado pela presença de variação de escala, uma vez que o conjunto de dados contém imagens adquiridas com níveis de ampliação de 100× e 400×. Os resultados quantitativos obtidos estão descritos na Tabela 3. Na segunda etapa, considerando apenas o modelo A-EKD, foram feitos testes utilizando imagens com resolução 100x e 400x e os respectivos resultados estão na Tabela 4.

A avaliação experimental foi conduzida com o objetivo de investigar a eficácia do mecanismo A-EKD frente aos desafios impostos pela variação de escala e pela incerteza epistêmica. Os resultados obtidos são comparados com quatro abordagens de referência, que foram replicadas seguindo a metodologia descrita nos artigos de cada autor: a destilação clássica (Vanilla Knowledge Distillation) proposta por Hinton et al. [Hinton et al. 2015], o método estado da arte em destilação desacoplada (Decoupled Knowledge Distillation – DKD) apresentado por Zhao et al. [Zhao et al. 2022], e o modelo convolucional–transformer desenvolvido por Nogueira et al. [Nogueira and Gomes 2025]. Adicionalmente, os resultados do método proposto são comparados com o modelo EKD evidencial descrito por Xiang et al. [Xiang et al. 2025].

Tabela 3. Desempenho dos modelos para classificação de tecidos de OSCC

Método	Acurácia (%)	F1-Score	ECE	Err-AUC
Conv+Transf. [Nogueira and Gomes 2025]	87,70	0,8540	–	–
Vanilla KD [Hinton et al. 2015]	81,15	0,7926	0,2771	0,8488
DKD (CVPR '22) [Zhao et al. 2022]	82,38	0,8308	0,3007	0,7667
EKD (SOTA '25) [Xiang et al. 2025]	86,48	0,8534	0,3312	0,6835
A-EKD (Proposto)	**90,98**	**0,9087**	**0,2627**	**0,8490**

Diante dos resultados da Tabela 3 observa-se que o método tradicional Vanilla KD apresentou acurácia inferior aos demais que os demais modelos (81,15%). Este era um resultado esperado, tendo em vista a maior simplicidade do modelo, devido a complexidade de cenários quando o Professor apresenta alta entropia global. O método A-EKD obteve desempenho acima dos demais, alcançando 90,98%, superando o desempenho de [Nogueira and Gomes 2025] ($90,98\%$ vs $87,70\%$).

O cenário da Tabela 3, representa o desafio mais próximo da realidade clínica de triagem, em que o sistema deve ser robusto a variações de escala, artefatos de digitalização e heterogeneidade tecidual. Observa-se que métodos de destilação considerados o estado da arte para imagens naturais, como o DKD e Vanilla KD, falharam em superar o baseline de $87,70\%$. Isso sugere a presença de "conhecimento corrompido" que foi destilado do Professor para o modelo Aluno. A heterogeneidade das escalas pode ter induzido o modelo Professor a gerar distribuições de incerteza ruidosas, degradando a generalização do Aluno.

A abordagem proposta A-EKD foi a única capaz de reverter essa tendência, atingindo uma acurácia de $90,98\%$. A técnica junção da técnica de *evidencial learning* com a destilação clássica aparenta ter atuado efetivamente como um filtro de qualidade, permitindo que o sistema ponderasse o Professor nas instâncias de confusão e aproveitasse o aprendizado supervisionado, resultando em um ganho líquido de desempenho.

O método proposto alcançou uma acurácia de $90,98\%$, representando um salto de $+4,5\%$ em relação ao EKD [Xiang et al. 2025] e $+8,6\%$ sobre o DKD [Zhao et al. 2022]. Mais relevante para o contexto clínico de OSCC é o F1-Score de $0,9087$. Enquanto o DKD foca na decomposição de *logits* entre classes alvo e não-alvo para melhorar a transferência de conhecimento [Zhao et al. 2022], ele falha em capturar a incerteza estrutural, o que limita sua capacidade de generalização em imagens histopatológicas complexas. O ganho no F1-Score do A-EKD valida que o métrica adaptativa não apenas melhora a precisão, mas equilibra a sensibilidade necessária para o diagnóstico oncológico.

Um dos achados mais impactantes deste experimento é a redução do ECE. O método EKD convencional [Xiang et al. 2025], embora superior em acurácia ao Vanilla KD, apresentou o maior erro de calibração ($0,3312$), indicando uma tendência à superconfiança herdada do Professor sem filtragem. Em contrapartida, o A-EKD obteve o menor ECE da tabela ($0,2627$). Isso mostra que a métrica adaptativa proposta ($\mathcal{G}$) atua como um regulador de calibração: ao ignorar o Professor em momentos de alta vacuidade ($u^T \to 1$), o Aluno (MobileNetV3) evita a absorção de incertezas. Este resultado corrobora a crítica [Zhang et al. 2025] sobre o "Efeito Tóxico" do Professor.

O valor de Err-AUC $(0,8490)$ do A-EKD, alinhado ao Vanilla KD mas com acurácia superior, sugere que o modelo mantém uma excelente capacidade de distinguir entre acertos e erros. O baixo desempenho do EKD estático neste quesito $(0,6835)$ reforça a hipótese de que a destilação de segunda ordem sem a métrica adaptativa distorce a fronteira de decisão do Aluno em amostras ambíguas. O A-EKD consegue o unir ambos pontos positivos: a precisão de um modelo profundo e a honestidade de um modelo calibrado.

Também foram investigadas as imagens separadas por resolução (100x e 400x) e os resultados para esse experimento estão dispostos na Tabela 4.

Tabela 4. Comparação de desempenho do método A-EKD em diferentes resoluções de imagem.

Resolução	ACC (%)	F1-Score	ECE ↓	Err-AUC ↑
100x	**94,23**	**0,9436**	0,2964	**0,8537**
400x	89,21	0,8917	**0,2596**	0,8113
Ambas (100x + 400x)	90,98	0,9087	0,2627	0,8490

A avaliação do impacto da escala de ampliação no diagnóstico de OSCC revelou variações significativas na relação entre precisão e calibração. O modelo treinado exclusivamente com imagens de 100x obteve o desempenho preditivo superior, com $94,23\%$ de acurácia, sugerindo que os padrões macroarquiteturais do epitélio oral são altamente discriminativos para a rede.

Por outro lado, as amostras de 400x, que focam em detalhes citológicos e nucleares, resultaram no menor erro de calibração (ECE de $0,2596$), indicando que a alta resolução fornece pistas fundamentais para que o modelo quantifique sua incerteza de forma mais precisa. O uso simultâneo de ambas as resoluções demonstrou um equilíbrio robusto, mantendo uma acurácia elevada $(90,98\%)$ e uma forte capacidade de detecção de erro (Err-AUC de $0,8490$). Este comportamento valida a premissa de que a fusão de escalas permite ao modelo A-EKD capturar tanto a desorganização tecidual (escala macro) quanto a atipia celular (escala micro), mitigando os riscos de superconfiança destacados na literatura atual.

Para avaliar a confiança do modelo, foram realizados testes estáticos onde foi considerado o conjunto de dados com resolução multipla (100x e 400x). A escolha desse *dataset* teve como premissa a maior proximidade da situação real em uma clinica. Os resultados estão descritos na Tabela 5.

Tabela 5. Resultados de confiabilidade estatística do modelo A-EKD

Métrica	Valor Obtido (A-EKD)
Intervalo de Confiança (IC 95%)	$\pm$ 6,44%
Brier Score	0,1342
Negative Log-Likelihood (NLL)	0,4409
Teste de McNemar (p)	$< 0,001$
Odds Ratio (Tamanho do Efeito)	24,5

A análise da confiabilidade estatística e calibração do modelo A-EKD demons-

tra uma predição confiável para o diagnóstico de carcinoma oral. Embora o Intervalo de Confiança (IC 95%) de $6,44\%$ reflita a limitação do tamanho da amostra de teste ($N = 116$) estabelecida segundo o protocolo de [Nogueira and Gomes 2025], a qualidade das predições é embasada pelo *Brier Score* de $0,1342$, além do *Negative Log-Likelihood* (NLL) de $0,4409$, comprovam um ajuste probabilístico preciso e a minimização de erros superconfiantes. A significância desses resultados é validada pelo Teste de McNemar com $p < 0,001$, o que rejeita a hipótese nula de igualdade entre os modelos e prova que a superioridade do aluno sobre o professor é estatisticamente sólida e não fruto do acaso. Complementarmente, o expressivo *Odds Ratio* de $24,5$ indica que o A-EKD possui uma probabilidade $24,5$ vezes maior de obter o diagnóstico correto em casos onde o modelo professor falharia, consolidando a eficácia do *Gate* na filtragem de incertezas e na transferência de conhecimento crítico para a patologia digital

5. Conclusão

A metodologia proposta por este trabalho, mostrou ser capaz de quantificar a qualidade da supervisão fornecida pelo modelo Professor e reduzir a transferência de conhecimento corrompido. Enquanto métodos do estado da arte como DKD e EKD sofreram degradação de desempenho (caindo para $82,38\%$ e $86,48\%$, respectivamente), o modelo A-EKD atingiu $90,98\%$ de acurácia, superando até mesmo arquiteturas complexas (EfficientNet + Transformer) reportadas na literatura recente ($87,70\%$).

Apesar dos avanços alcançados, este estudo abre novas frentes de investigação que delineiam os próximos passos para a evolução desta tecnologia. A atual abordagem limita-se à classificação binária (Normal vs. OSCC). Um passo natural é estender para uso de redes de segmentação (como U-Net), permitindo não apenas detectar a doença, mas delimitar as margens da lesão para auxiliar em procedimentos cirúrgicos. O diagnóstico clínico exige a diferenciação entre graus de displasia (leve, moderada, severa) e carcinoma in situ. Trabalhos futuros devem validar o comportamento do *Gate* em cenários de classificação multiclasse, onde a incerteza de fronteira é ainda mais crítica. Além disso, testes envolvendo *split patient-wise* e execuções independentes serão analisados.

6. Agradecimentos

Agradeço à FAPEMIG pela contribuição através do apoio financeiro (Bolsa de pesquisa concedida) que tornou este trabalho possível e ao Programa de Pós-Graduação em Ciência da Computação (PPGCO) da FACOM/UFU.

Referências

Agresti, A. (2012). *Categorical Data Analysis*, volume 792. John Wiley & Sons, 3rd edition.

Brier, G. W. (1950). Verification of forecasts expressed in terms of probability. *Monthly Weather Review*, 78(1):1–3.

Guo, C., Pleiss, G., Sun, Y., and Weinberger, K. Q. (2017). On calibration of modern neural networks. In *International Conference on Machine Learning (ICML)*, pages 1321–1330.

Hendrycks, D. and Gimpel, K. (2017). A baseline for detecting misclassified and out-of-distribution examples in neural networks. *International Conference on Learning Representations (ICLR)*.

Hinton, G., Vinyals, O., and Dean, J. (2015). Distilling the knowledge in a neural network. *arXiv preprint arXiv:1503.02531*.

Howard, A. et al. (2019). Searching for mobilenetv3. In *Proceedings of the IEEE/CVF International Conference on Computer Vision*, pages 1314–1324.

Lambert, B., Forbes, F., Doyle, S., H., D., and Dojat, M. (2024). Trustworthy clinical ai solutions: A unified review of uncertainty quantification in deep learning models for medical image analysis. *Artificial Intelligence in Medicine*, 150:102830.

McNemar, Q. (1947). Note on the sampling error of the difference between correlated proportions or percentages. *Psychometrika*, 12(2):153–157.

Nogueira, M. and Gomes, E. F. (2025). Histopathological imaging dataset for oral cancer analysis: A study with a data leakage warning. In *Proceedings of the 18th International Joint Conference on Biomedical Engineering Systems and Technologies*, pages 811–818.

Paraíso, E. and Machado, A. (2025). Impacto do balanceamento e regularização na segmentação semântica de imagens histopatológicas. In *Anais Estendidos do XXV Simpósio Brasileiro de Computação Aplicada à Saúde*, pages 13–18, Porto Alegre, RS, Brasil. SBC.

Prajwal, R., Pawan, S. J., Nazarian, S., Heller, N., Weight, C. J., Duddalwar, V., and Kuo, C.-C. J. (2025). A study on energy consumption in ai-driven medical image segmentation. *Journal of Imaging*, 11(6).

Rahman, T. Y., Mahanta, L. B., Das, A. K., and Sarma, J. D. (2020). Histopathological imaging database for oral cancer analysis. *Data in Brief*, 29:105114.

Sensoy, M., Kaplan, L., and Kandemir, M. (2018). Evidential deep learning to quantify classification uncertainty. In *Advances in Neural Information Processing Systems (NeurIPS)*, volume 31.

Tan, M. and Le, Q. (2019). Efficientnet: Rethinking model scaling for convolutional neural networks. *arXiv preprint arXiv:1905.11946*.

Van Rijsbergen, C. (1975). *Information Retrieval*. Butterworths.

WHO Classification of Tumours Editorial Board, editor (2024). *Head and neck tumours*, volume 9 of *World Health Organization classification of tumours*. IARC, Lyon, 5 edition.

Xiang, L., Gao, J., and Xu, C. (2025). Evidential knowledge distillation. In *Proceedings of the IEEE/CVF International Conference on Computer Vision (ICCV)*.

Zhang, J., Gao, Y., Liu, R., Cheng, X., Zhang, H., and Chen, S. (2025). Can students beyond the teacher? distilling knowledge from teacher's bias. In *Proceedings of the AAAI Conference on Artificial Intelligence*, AAAI'25/IAAI'25/EAAI'25. AAAI Press.

Zhao, B. et al. (2022). Decoupled knowledge distillation. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pages 11953–11962.