

**DETECÇÃO DE PSORÍASE VULGAR EM PACIENTES BRASILEIROS
POR MEIO DE UM SISTEMA COMPUTER-AIDED DIAGNOSIS
BASEADO EM VISUAL TRANSFORMERS**

Lucas Gargalhoni Antunes Corrêa¹

Apoio: PIVIC Mackenzie

Gustavo Scalabrini Sampaio¹

Faculdade de Computação e Informática (FCI)

Universidade Presbiteriana Mackenzie

São Paulo, SP – Brasil

<lucasgargalhoni@mackenzista.com.br>

<gustavo.sampaio@mackenzie.br>

Resumo. A psoríase é uma doença dermatológica crônica que afeta mais de 5 milhões de brasileiros, representando um desafio significativo para o sistema de saúde nacional. Este trabalho apresenta o desenvolvimento de um sistema de diagnóstico assistido por computador (CAD) baseado em Visual Transformers para classificação automática de psoríase vulgar e dermatite utilizando imagens de pacientes brasileiros. A metodologia envolveu a implementação de um pipeline completo incluindo coleta e curadoria de dados fornecidos pela Faculdade de Medicina do ABC (FMABC), pré-processamento com técnicas de remoção de fundo e data augmentation, e treinamento de um modelo baseado na arquitetura Swin Transformers. A validação foi realizada através de validação cruzada k-fold estratificada, e a interpretabilidade do modelo foi analisada usando Grad-CAM. Os resultados demonstraram desempenho promissor e confirmaram que o modelo focaliza adequadamente nas características morfológicas clinicamente relevantes das lesões. A abordagem proposta apresenta potencial significativo como ferramenta de apoio ao diagnóstico dermatológico em cenários clínicos brasileiros.

Palavras-chave: *Psoríase, Diagnóstico Assistido por Computador, Visual Transformers, Swin Transformers, Dermatologia, Inteligência Artificial, Processamento de Imagens Médicas*

Abstract. Psoriasis is a chronic dermatological disease affecting over 5 million Brazilians, representing a significant challenge for the national healthcare system. This work presents the development of a Computer-Aided Diagnosis (CAD) system based on Visual Transformers for automatic classification of psoriasis vulgaris and dermatitis using images from Brazilian patients. The methodology involved implementing a complete pipeline including data collection and curation provided by the ABC Medical School (FMABC), preprocessing with background removal techniques and data augmentation, and training a model based on the Swin Transformers architecture. Validation was performed through stratified k-fold cross-validation, and model interpretability was analyzed using Grad-CAM. Results demonstrated promising performance and confirmed that the model adequately focuses on clinically relevant morphological characteristics of lesions. The proposed approach shows significant potential as a diagnostic support tool in Brazilian clinical dermatology scenarios.

Keywords: *Psoriasis, Computer-Aided Diagnosis, Visual Transformers, Swin Transformers, Dermatology, Artificial Intelligence, Medical Image Processing*

1 Introdução

A relação entre a computação e área da saúde está cada vez mais próxima, prova disso são as diversas aplicações da computação na medicina, desde plataformas de atendimento remoto (telemedicina) até sistemas de diagnóstico guiado por computador (Computer-Aided Diagnosis (CAD)). Os sistemas CAD podem ser compostos de diferentes técnicas da computação. Em pesquisas recentes, esses sistemas têm sido beneficiados pelo uso da IA, uma área da computação que tem a intenção de replicar a inteligência humana. Uma subárea da IA importante para os sistemas CAD é a Visão Computacional, cuja meta é utilizar computadores para emular a visão humana, incluindo o aprendizado e a capacidade de fazer inferências e agir com base em informações visuais. Uma área que age em conjunto com a Visão Computacional é o Processamento Digital de Imagens, que geralmente está atribuída no estudo e aplicação de técnicas de mais baixo nível, como o realce de contraste e aguçamento de imagens (GONZALEZ; WOODS, 2008).

Na dermatologia, existe uma lacuna entre pacientes de doenças de pele e a expertise necessária para lidar com eles (HAMEED et al., 2019). Pessoas que vivem nas áreas rurais são as que mais se prejudicam por conta da falta de recursos, aponta pesquisa da Organização Mundial de Saúde (2015). Os sistemas CAD são muito vantajosos nesses cenários, oferecendo um pré-diagnóstico de diversas doenças. Além disso, esses sistemas buscam proporcionar ao processo de definição da doença maior precisão, garantindo o tratamento adequado ao paciente e diminuindo custos operacionais ((HAMEED et al., 2019), (DASH et al., 2020), (ARORA et al., 2021). Uma doença dermatológica que pode ser analisada e diagnosticada por sistemas CAD é a psoríase. A psoríase é uma doença inflamatória, crônica e recorrente que, de acordo com a Sociedade Brasileira de Dermatologia (2021), afeta a pele e articulações de mais de 5 milhões de brasileiros.

Existem vários fenótipos dessa doença, sendo a mais comum a psoríase vulgar, presente em cerca de 90% dos casos (GRIFFITHS; BARKER, 2007). Essa doença não apresenta risco à vida diretamente, porém traz diversas outras implicações para o paciente, desde coceira até sangramento na região lesionada. Estudos epidemiológicos identificaram também uma alta prevalência de fatores de risco cardiovascular em pacientes psoriáticos.

Essa característica é particularmente importante porque, as doenças cardiovasculares – incluindo a síndrome metabólica, obesidade, hipertensão, *diabetes mellitus*, resistência à insulina e a dislipidemia (MILLER et al., 2013) – são a principal causa de morte, hospitalizações e atendimentos ambulatoriais em todo o mundo, inclusive em países em desenvolvimento como o Brasil (BARROSO et al., 2021).

A psoríase tem seu diagnóstico realizado inicialmente através de uma análise visual, ou seja, pela aparência clínica e distribuição da lesão. Identificada, ela é classificada como leve, moderada ou grave, sendo essa classificação baseada principalmente na superfície corporal afetada e no efeito da lesão no paciente. Diante disso, uma medida para a classificação da doença é a Psoriasis Area and Severity Index (PASI), que é dividida em dois passos, o primeiro é o cálculo da Área de superfície corporal (BSA) afetada com as lesões e o segundo passo consiste em avaliar o eritema (vermelhidão na pele), o endurecimento (espessura) e a descamação das lesões (Sociedade Brasileira de Dermatologia, 2018).

1.1 Objetivos

O presente projeto tem como objetivo conduzir um estudo motivado pelo alto número de pacientes afetados pela psoríase, e também oportunidades de utilização de dados de pacientes brasileiros, para desenvolver um sistema CAD dermatológico, baseado em um modelo de IA utilizando técnicas de visão computacional, processamento digital de imagens para diagnóstico de psoríase vulgar. Este trabalho dará início ao desenvolvimento de um sistema CAD próprio, utilizando imagens e dados clínicos de pacientes brasileiros. Essa característica é um dos pontos chaves do estudo, já que devido a fatores como a região, clima e exposição solar, a amostragem local é muito diferente comparada a outros trabalhos já propostos. Para atingir essa meta foram estipulados os seguintes objetivos específicos:

- Definir e coletar as bases para o treinamento do modelo;
- Definir a arquitetura e treinar um modelo de IA para a classificação entre psoríase vulgar e dermatite; usando imagens dermatológicas;
- Validar os resultados de desempenho do modelo treinado;
- Analisar a viabilidade técnica de aplicação do método proposto em um produto real.

2 Referencial teórico

Este projeto de pesquisa tem como base de estudo três grandes áreas: sistemas de diagnóstico assistido por computador, redes neurais convolucionais e processamento digital de imagens. De forma individual, cada uma dessas áreas pode ter aplicações em diferentes setores, porém, com os avanços e a popularidade do uso de técnicas de Inteligência Artificial (IA) na saúde, a relação entre essas áreas de estudo está cada vez mais forte. Essa seção está dividida entre as áreas de estudo e aplicação do projeto, sendo inteligência artificial e sistemas CAD respectivamente.

2.1 Redes Neurais Artificiais

Segundo Russell e Norvig (2010), a Inteligência Artificial é um campo da computação que estuda a construção de entidades inteligentes, ou seja, máquinas que parecem ter inteligência humana. A construção dessas entidades se beneficiou de diversas áreas de estudo, como por exemplo: filosofia, matemática, economia, neurociência, psicologia, engenharia da computação, linguística, teoria do controle, cibernética e linguística.

Dentro do campo de estudo da inteligência artificial, uma subárea muito importante é o aprendizado de máquina. Aprendizado de máquina envolve o estudo e aplicação de algoritmos que possam aprender a partir de dados, esses algoritmos aprimoram seu desempenho em uma tarefa à medida que são expostos a mais dados. Dentro dessa área de estudo existem três tipos de aprendizados, aprendizado não supervisionado, em que durante o aprendizado o agente aprende padrões de uma entrada mesmo sem ter recebido valores de saída explícitos; aprendizado por reforço, no qual o sistema aprende a partir de uma série de reforços, que são sinais para as decisões do agente, sejam negativos caso a decisão seja ruim, ou positivo em uma boa decisão; e por último, um tipo de aprendizado muito popular para o treinamento de sistemas CAD, o aprendizado supervisionado, nesse tipo de aprendizado, o sistema recebe um conjunto de dados e suas respostas, e ao decorrer das iterações o modelo aprende a função que mapeia a entrada com a saída (HAYKIN, 2009).

As técnicas de aprendizado de máquina tiveram grande impacto neste estudo, possibilitando o treinamento do sistema proposto. Mais especificamente, foi implementado aprendizado profundo (*Deep learning* (DL)), esse campo concentra uma ampla família de técnicas de aprendizado de máquina em que as hipóteses assumem a forma de circuitos

algébricos complexos com forças de conexão ajustáveis. O termo ‘profundo’ refere-se ao fato de que esses sistemas geralmente estão organizados em camadas (RUSSELL; NORVIG, 2009).

Os sistemas baseados em redes neurais surgiram a partir do desafio de resolver problemas complexos que são inviáveis de solucionar com paradigmas tradicionais de programação. Esses sistemas são inspirados no funcionamento do neurônio humano, nesse caso um neurônio artificial, também conhecido como *perceptron*.

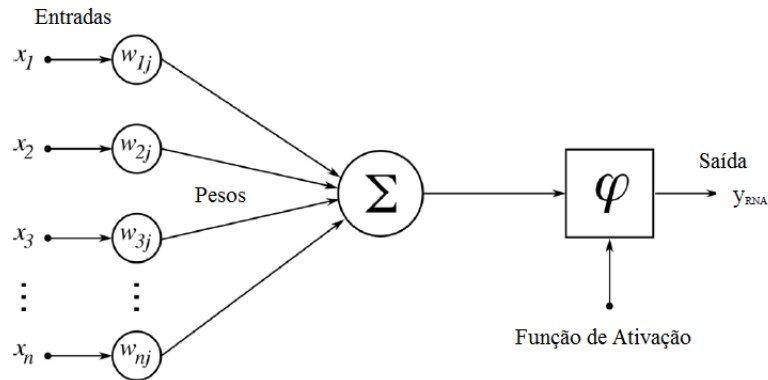


Figura 1 – Exemplo do modelo perceptron (MARTINIANO et al., 2016).

O modelo do perceptron, demonstrado na figura 1, é composto por algumas características principais: Um número n de entradas, em que cada entrada será ponderada por seu respectivo peso w , e um termo de viés b (bias). O funcionamento desse neurônio se inicia com uma combinação linear, que é a soma ponderada das entradas x e dos pesos w , essa operação pode ser representada como: $z = w_1x_1 + \dots + w_nx_n$. O resultado dessa operação (z) é passado então a uma função de ativação. Essa função de ativação tem a capacidade de amplificar o aprendizado desses sistemas, já que é essa função que determina se o neurônio será ativado ou não. Essa função de ativação é responsável por aplicar uma transformação não linear à saída da primeira operação, resultando na operação final representada por $\sigma = f(z + b)$, em que b é termo de viés e z o resultado do somatório (HAYKIN, 2009).

Considerando o entendimento do perceptron, a arquitetura baseada em redes neurais é dividida em basicamente três partes: A camada inicial, essa camada é responsável por receber os dados de entrada; as camadas ocultas, conhecidas também como *hidden layers*, em que as informações são processadas, e a camada de saída, em que é gerado o valor de saída da rede. Além das redes neurais serem compostas por camadas de neurônios, cada neurônio em uma camada está conectado com todos os neurônios da

camada seguinte, logo temos a formação de uma rede *Multilayer Perceptron* (MLP). Para análise de imagens, uma arquitetura que tem se mostrado muito eficiente é a rede neural convolucional (Convolutional Neural Network (CNN)), essa arquitetura segue a mesma lógica dos modelos baseados em redes neurais, porém com um adicional, as convoluções.

A convolução é um processo de filtragem espacial (plano que contém os pixels da imagem), que consiste em aplicar o somatório do produto entre duas matrizes, a imagem e uma máscara ao longo da região que estas se sobrepõem, sendo a imagem uma função bidimensional $f(i, j)$, em que i e j são as coordenadas, e a amplitude de f em qualquer par de coordenadas se refere a intensidade de cor naquele ponto, já a máscara uma matriz de tamanho variado (GONZALEZ; WOODS, 2008).

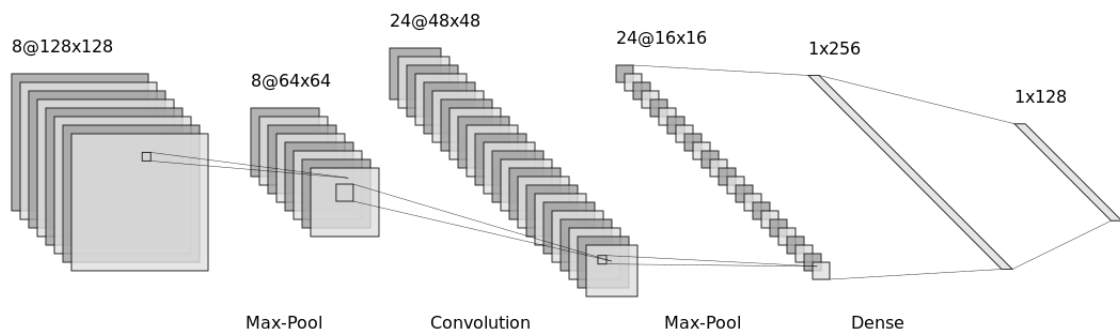


Figura 2 – Exemplo rede neural convolucional.

As redes convolucionais, que seguem o modelo apresentado na figura 2, geralmente recebem como entrada imagens e podem fazer extração de características e classificações a partir desses dados. Em casos de diagnóstico médico feito a partir da avaliação visual, como uma parte do diagnóstico da psoríase, por exemplo, o uso desses modelos seria adequado.

Uma alternativa à abordagem da CNN é a arquitetura de *Visual Transformers*. Essa arquitetura utiliza o algoritmo de *Transformers* (figura 3) para análise de imagens, os *transformers*, foram inicialmente propostos por Vaswani et al. (2017). Diferente dos modelos recorrentes da época, que processavam informações de forma sequencial, o que impedia a paralelização eficiente dentro dos exemplos de treinamento em sequências longas e limitava o agrupamento de dados, o *Transformer* abandona a recorrência. Em vez disso, ele se baseia inteiramente em um mecanismo de atenção para capturar dependências globais entre as entradas e saídas. Essa abordagem permitiu uma paralelização significativamente

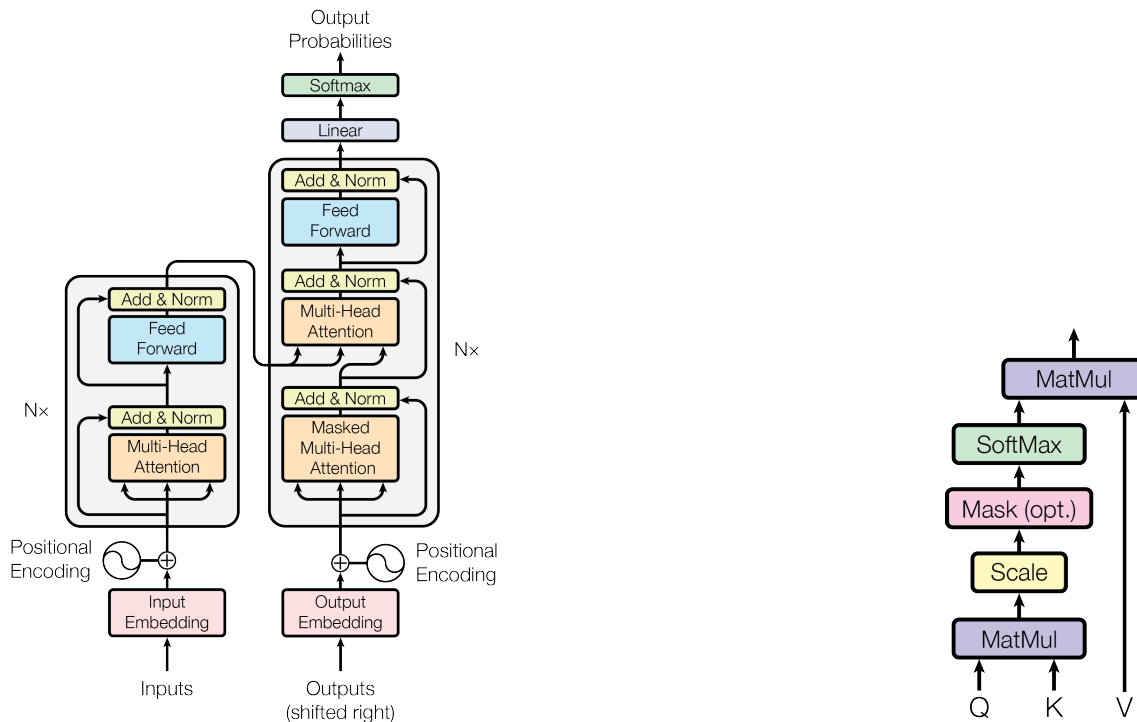


Figura 3 – Arquitetura do transformer. (Vaswani et al. (2017))

Figura 4 – Scaled Dot-Product Attention. (Vaswani et al. (2017))

maior durante o treinamento e estabeleceu novos patamares de desempenho em tarefas como a tradução automática.

Vaswani et al. (2017) define o mecanismo de autoatenção (*self-attention*), demonstrado na figura 4, como um algoritmo que permite que o modelo pese a importância de diferentes partes da sequência de entrada ao gerar uma representação para cada elemento da sequência. As entradas são recebidas como vetores. Ao processar esse vetor, o modelo compara diretamente com todos os outros vetores na sequência de entrada simultaneamente em uma única camada, o que possibilita a paralelização, e determina quais são as mais relevantes para entender o contexto daquele elemento específico, diferentemente da CNN, que consideram apenas as informações que se encontram no mesmo espaço do filtro definido, ou seja, contexto local. Esse processo é realizado calculando scores de similaridade entre os elementos, permitindo que as dependências de longo alcance sejam capturadas de forma eficaz, sem as limitações sequenciais dos modelos recorrentes.

Os *Visual Transformers* (figura 5), por sua vez, dividem as imagens em patches de tamanhos fixos, diferentemente das CNNs que processam pixels diretamente ou por meio de filtros convolucionais. Cada um desses patches é vetorizado e a esses vetores são adicionados embutimentos de posição (*position embeddings*), para preservar a informação espacial original da imagem, após essa operação, essa sequência passa por um *encoder*

padão que utiliza o algoritmo de atenção para aprender os padrões daquela sequência e as relações entre os diferentes patches da imagem. Para realizar tarefas de classificação, como a detecção de doenças de pele, um token de classificação extra, que é treinável, é adicionado à sequência, e sua saída final é usada para a previsão (DOSOVITSKIY et al., 2020).

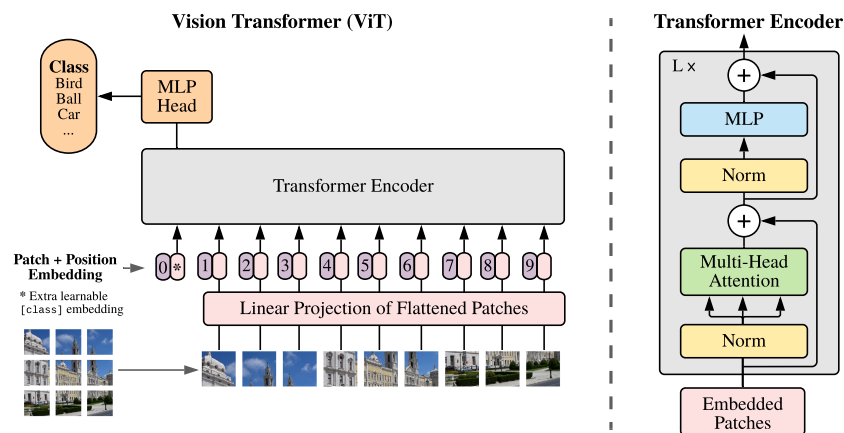


Figura 5 – The Visual Transformer. (Dosovitskiy et al. (2020))

2.2 Sistemas CAD

Os primeiros estudos sobre análise quantitativa de imagens médicas por computador foram relatados na década de 1960. Porém, nos anos 80 surgiu uma nova abordagem, que assume que os sistemas CAD podem ser utilizados pelos profissionais da saúde como ferramenta e não para substituí-los. Assim, o objetivo principal desses sistemas é que a saída do CAD seja como uma “segunda opinião” para os médicos, para que estes façam a avaliação final. Nesse sentido, o uso dos sistemas CAD permite uma melhora na avaliação final em diferentes cenários, caso o médico esteja menos confiante sobre sua avaliação, a saída desse sistema tem a intenção de apoiar a avaliação final; e em uma situação onde o médico está mais confiante de sua avaliação, ele pode “descartar” a saída do computador. Alguns estudos afirmam que os sistemas CAD não precisam necessariamente se igualar ou superar a precisão na classificação em comparação aos profissionais na área da saúde, porém é importante ressaltar que quanto maior for o desempenho da máquina, melhor será o resultado final combinado (DOI, 2007).

Para auxiliar na definição da metodologia do projeto de pesquisa, foi realizado um levantamento do estado da arte para as técnicas de classificação com imagens médicas. O processo de levantamento bibliográfico considerou o levantamento de artigos científicos nas bases científicas MDPI, IEEE, Elsevier e Scopus. Para realizar a busca dos artigos, foram utilizadas as palavras-chaves: “CAD”, “Computer Aided System”, “CAD convolution”, “psoriasis”, “CAD psoriasis”. Foram levantados 21 artigos considerando o filtro por título e resumo; dos quais, 13 foram detalhadamente analisados, com destaque para os trabalhos de Li (2020), Brinker (2019) e Doi (2007).

Li (2020), apresentou em sua pesquisa um sistema CAD para o diagnóstico de câncer de pâncreas. Nesse sistema, o autor utilizou um algoritmo baseado em support vector machine (SVM). Além desse algoritmo, o autor utilizou outras técnicas como o algoritmo LASSO, técnica de regularização usada principalmente para evitar o sobre ajuste do modelo e ajudar a selecionar as características importantes, e também propôs um modelo utilizando ensemble, que visa trazer mais robustez e melhorar a precisão do sistema através da construção e combinação de vários modelos. O algoritmo de ensemble usado para esse sistema foi o algoritmo bagging (bootstrap aggregating); nesse algoritmo várias subamostras são geradas aleatoriamente a partir do conjunto de dados original, com repetição e as previsões dos modelos individuais são combinadas para gerar uma previsão final. Esse sistema CAD atingiu uma precisão de 91.63% para os dados na classe “Normal Stage - IV”, essa classe representa o estágio mais avançado da doença.

Já Brinker et al. (2019), apresentou em sua pesquisa um sistema CAD voltado para dermatologia, o modelo proposto superou dermatologistas na classificação de imagens de melanoma. Em seu trabalho, Brinker utilizou apenas imagens com uma qualidade avaliada como excelente, boa ou suficiente pelos dermatologistas que participaram da pesquisa. O sistema CAD proposto foi desenvolvido usando uma arquitetura de redes neurais já conhecida, a ResNet50 e os dados vieram de fontes open-source (código aberto) e comprovados por biópsia. Considerando um intervalo de confiança (CI) de 95%, após o treinamento do modelo, este atingiu uma sensibilidade de 82.3% (95% CI: 78.3–85.7%) e especificidade de 77.9% (95% CI: 73.8–81.8%), já os dermatologistas atingiram uma sensibilidade de 67.2% (95% CI: 62.6–71.1%) e especificidade de 62.2% (95% CI: 57.6–66.9%).

Ainda no contexto dermatológico, DASH et al. apresenta um sistema CAD totalmente integrado, que é composto por três estágios: classificação, segmentação e avaliação

da severidade da lesão. No primeiro estágio, o sistema é baseado numa arquitetura de redes convolucionais chamada de VGG-16, composta de 13 camadas convolucionais e 3 camadas totalmente conectadas. Neste trabalho, algumas alterações foram feitas nesse modelo, como a remoção de camadas “redundantes”, com o objetivo de reduzir a quantidade de parâmetros, tornando assim essa rede mais eficiente computacionalmente. A primeira parte do trabalho, que compõe a rede convolucional para classificação da doença, atingiu uma acurácia de 93.76%.

Como já visto anteriormente, em sua grande maioria, esses sistemas são construídos para lidar com dados não estruturados, que é o caso de imagens. A visão humana, diferente dos computadores, é limitada à banda visual do EM, já os aparelhos de processamento de imagens cobrem quase todo o espectro eletromagnético (EM), variando de ondas gama a ondas de rádio; basicamente esses dispositivos são capazes de trabalhar com imagens geradas por fontes que os seres humanos não estão acostumados a associar, entre esses tipos de fonte, estão: imagens geradas por ultrassons, microscopia eletrônica e imagens gerada por computador (GONZALEZ; WOODS, 2008).

Há uma grande vantagem na aplicação da visão computacional nesse setor, que por sua vez, estuda diferentes técnicas para trabalhar com imagens, como por exemplo, os já citados modelos de classificação, além de técnicas de extração de características, detecção de bordas, entre outras. Entre essas diversas técnicas, destacam-se os modelos baseados na arquitetura de redes neurais (DOI, 2007).

Do mesmo modo que a visão computacional avança com suas técnicas na área da saúde, outra área que também se beneficia é o processamento digital de imagens, que possui uma relação muito forte com a visão computacional. Essa área de estudo está ligada principalmente com os conceitos de filtragem, transformações geométricas e restauração de imagens. O objetivo desses tratamentos nas imagens é prepará-las para análises posteriores, de modo a oferecer as técnicas necessárias para pré-processar e manipular imagens, antes de “alimentar” esses dados nas redes convolucionais (GONZALEZ; WOODS, 2008).

3 Metodologia

O presente projeto apresenta o desenvolvimento de um sistema CAD, a partir de uma arquitetura de *visual transformers*. Esse modelo terá a tarefa de lidar com a base de dados de imagens que foi fornecida pela Faculdade de Medicina do ABC (FMABC), composta

de imagens de pacientes brasileiros afetados pela psoríase e pacientes com dermatite. A intenção dessa proposta é que a combinação de diferentes análises e informações traga um aumento na precisão desse tipo de sistema, fornecendo um suporte à tomada de decisão. Efetivamente, será implementada uma solução de ponta a ponta, que exigirá desde técnicas para o tratamento e preparo dos dados, até soluções para o refinamento do modelo de IA.

A partir dos conceitos de redes neurais artificiais e *transformers*, fica mais claro entender a construção dos modelos de IA, que seguem um certo fluxo padrão, esse fluxo é conhecido como *pipeline*. Os *pipelines* de IA envolvem todo o fluxo de atividades necessárias para o funcionamento apropriado do modelo. De modo genérico, este fluxo geralmente começa com atividades relacionadas à extração, transformação e carregamento dos dados, processo conhecido como *Extract Transform Load* (ETL); no caso de modelos de IA voltados a visão computacional, essa fase utiliza técnicas para preparar essas imagens a um modelo de IA, geralmente envolvendo atividades relacionadas ao processamento digital de imagens. Além das transformações nas imagens é necessário realizar uma estratificação dos dados, uma parte deve ser separada para o treinamento da rede, enquanto a outra deve ser exclusivamente voltada para o teste da rede. A etapa seguinte envolve o treinamento da rede, essa etapa exige diferentes técnicas para o sucesso do modelo, entre essas técnicas estão a escolha de um algoritmo de otimização adequado ao problema, o uso de uma função de perda apropriada, ajuste correto dos hiperparâmetros da rede no treinamento. As subseções a seguir destacam cada uma das fases do *pipeline* aplicado para o desenvolvimento do modelo.

3.1 Coleta de dados

A etapa de coleta de dados configura-se da responsabilidade de estabelecer formas consistentes de acesso aos dados fornecidos pela FMABC, também deve garantir segurança e a integridade dessas informações. Para a realização da coleta de dados, foram desenvolvidos *scripts* utilizando a linguagem de programação Go, uma ferramenta que ganhou muita popularidade devido ao seu desempenho. O desenvolvimento desses *scripts* permitiram que todo o manuseio das informações fosse realizado de forma automatizada, além de coletar os dados, também envolvem tarefas como catalogação das informações.



Figura 6 – Exemplo imagem de paciente com psoríase (FMABC, 2025)

3.1.1 Pré processamento

Durante essa fase, foram estudadas quais seriam as técnicas aplicadas para preparar os dados coletados ao modelo, dados estes exemplificado na figura 7. Nessa fase foram encontrados alguns desafios, entre eles o primeiro é a limitação no número de amostragens recebidas pela instituição parceira e desbalanceamento entre as classes de dados, ou seja, na prática haviam trinta vezes mais imagens da categoria psoríase.



Figura 7 – Imagem transformada de paciente com psoríase (FMABC, 2025)

Antes do uso desses dados para o modelo, foi necessário realizar uma inspeção manual sobre o dataset, que resultou na exclusão de 3000 imagens, devido a fatores como baixa qualidade, por exemplo imagens que estavam muito embaçadas ou com baixa resolução, e representação inadequada da classe, em que algumas fotografias que mostravam o corpo todo, e não a área lesionada. A utilização de tais imagens poderiam introduzir

ruído e prejudicar a capacidade de generalização do modelos.

Adicionalmente, para isolar a área de interesse, foi executado um pré-processamento específico para remover o fundo das imagens. Para este fim, foram desenvolvidos scripts customizados em *Golang* que aplicavam duas estratégias de segmentação. A primeira estratégia visava remover o fundo azul predominante em muitas imagens do dataset. O script identificava os pixels dentro de um intervalo de cor azul pré-definido e os substituía por uma cor neutra. A segunda, mais refinada, utilizava um filtro baseado em intervalos de cor para isolar a área da pele e remover o restante da imagem.

Para lidar com as limitações mencionadas anteriormente, foram desenvolvidos e aplicados algoritmos de *image augmentation*, esses algoritmos iteram sobre os dados fornecidos, e a partir destes, geram novos dados com filtros pré determinados. Neste trabalho, foi realizado uma única operação o *random cropping*, que aplica um corte numa posição aleatória da imagem.

Outro desafio encontrado durante o pré processamento foram as dimensões das imagens, já que as imagens tinham dimensões por volta de 2400x3800 pixels, e esse tipo de arquitetura trabalha com entradas de 521x521 pixels, assim todas as tentativas de aplicar redimensionamento nas imagens não tiveram bons resultados, pois houve uma perda muito grande de qualidade dos dados, a alternativa encontrada para lidar com isso foi o uso já mencionado de *random cropping*, pois o cropping descarta a necessidade de alterar os pixels da imagem, apenas usar um recorte de uma área aleatória.

No processo de *data augmentation*, também foi aplicado a rotação de 50º graus nas imagens e *flip* na horizontal e vertical, assim a mesma amostra porém em posição diferente poderia ser utilizada para treinamento da rede, o que enriqueceu o dataset e fortaleceu a capacidade de generalização do modelo. Além do aumento da base de dados, foram aplicadas transformações no domínio de cores das imagens. As imagens foram convertidas para o domínio de cor YCbCr, muitos *softwares* de edição de imagens utilizam esse domínio para manipulação desses dados. A principal vantagem desse domínio é a separação da informação de luminância (canal Y) da informação de crominância (canais Cb e Cr)

Por fim, todas as imagens foram normalizadas, um passo importante para o treinamento de redes neurais. O procedimento de normalização consistiu em ajustar os valores dos pixels para o intervalo $[0, 1]$, o que garante que a rede não seja afetada pela escala dos dados de entrada, promovendo uma convergência mais rápida e estável do

modelo.

3.2 Treinamento do modelo

Para avaliar o desempenho do modelo de forma robusta e imparcial, foi utilizada a validação cruzada com *k-folds* estratificada. O conjunto de dados foi dividido em *folds*, garantindo que a proporção das classes fosse mantida em cada partição de treinamento e teste. Essa abordagem minimizou a dependência do modelo em relação a uma única divisão de dados.

A partir de um processo exploratório, em que foram aplicados diferentes testes com modelos de CNN e *Visual Transformers*, para identificar a configuração mais adequada para a tarefa de classificação binária de doenças de pele. As arquiteturas investigadas incluíram modelos de redes neurais convolucionais (CNN), como VGG16 e ResNet152, e arquiteturas baseadas em transformadores, como Visual Transformers (*ViT*) e Swin Transformers. Após uma análise comparativa, a arquitetura *Swin Transformers* foi selecionada para o modelo final.

De modo a otimizar a busca por hiperparâmetros ideais para o objetivo proposto, foi utilizada uma técnica de *Grid Search*, que consiste em testar e avaliar o desempenho de todas as combinações de parâmetros previamente fornecidos de forma automática. A Tabela 1 apresenta os valores finais selecionados após a aplicação desta técnica.

A taxa de aprendizado foi explorada em um intervalo que incluiu valores como $[0.1, 0.01, 0.001]$, enquanto o decaimento de peso variou entre $[0.01, 0.001, 0.0001]$. O tamanho do lote também foi um parâmetro de otimização, testado com valores comuns na literatura como $[16, 32, 64]$. O valor final para o momento (*momentum*) foi ajustado com base nos valores mais promissores, com o valor de 0.9 resultando no melhor desempenho. Após a avaliação exaustiva de todas as combinações, o conjunto de valores que obteve o melhor desempenho no conjunto de validação foi selecionado como a configuração final para o treinamento do modelo, conforme detalhado na Tabela 1.

3.3 Processo de análise dos resultados

A avaliação do modelo incluiu análise quantitativa e qualitativa para compreender sua capacidade de tomada de decisão. Para tal, foram empregadas duas abordagens complementares: a análise de métricas de desempenho estatísticas e a interpretação visual por meio de mapas de calor.

Hiperparâmetro	Valor
Otimizador	SGD (<i>Stochastic Gradient Descent</i>)
Taxa de aprendizado	0.01
Momento (<i>Momentum</i>)	0.9
Dcaimento de peso	0.001
Função de perda	Entropia Cruzada Binária (BCELoss)
Tamanho do lote	32
Número de épocas	40
Scheduler	ReduceLROnPlateau

Tabela 1 – Hiperparâmetros finais utilizados no treinamento do modelo.

O desempenho do modelo foi mensurado utilizando métricas de classificação padrão para problemas binários. Conforme os resultados desses métodos, foi possível analisar se o desempenho do modelo está de acordo com o esperado. Dentre as técnicas mais comuns, foram utilizadas métricas como acurácia, definida pela fórmula

$$acurácia = \frac{VP + VN}{VP + FN + VN + FP} , \quad (1)$$

em que VP é a quantidade de verdadeiros positivos; VN é a quantidade de verdadeiros negativos; FN é a quantidade de falsos negativos e FP é a quantidade de falsos positivos. Em seguida, a métrica de precisão foi utilizada, determinada pela fórmula:

$$precisão = \frac{VP}{VP + FN} , \quad (2)$$

também, uma métrica de avaliação chamada sensibilidade, estabelecida pela fórmula:

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

foi empregada a revocação (*recall*), uma métrica que mede a capacidade do modelo de encontrar todas as amostras positivas, estabelecida pela fórmula:

$$revocação = \frac{VP}{VP + FN} \quad (4)$$

Por fim, o *F1-Score*, que representa a média harmônica entre precisão e revocação, foi utilizado como uma métrica de equilíbrio para avaliar o modelo em cenários com classes desbalanceadas.

$$F1 = 2 \cdot \frac{Precisão \cdot Revocação}{Precisão + Revocação} \quad (5)$$

Para estabelecer os mapas de calor e promover uma compreensão visual dos resultados do modelo, a ferramenta utilizada para essa tarefa foi o *GradCam*, uma ferramenta de interpretabilidade que gera mapas de calor para qualquer camada convolucional de uma rede neural. Esses mapas destacam as regiões da imagem de entrada que mais influenciaram a decisão do modelo para uma determinada classe.

4 Resultados e Discussão

A rede treinada apresentou resultados relevantes e muito promissores. Um dos principais desafios encontrado durante o processo de desenvolvimento foi a complexidade de trabalhar com o dataset, além do pré-processamento, também foi necessário a remoção de certas imagens que não estavam adequadas para o treinamento da rede, devido a diversos fatores externos, como por exemplo qualidade da foto. Mesmo em um ambiente controlado não havia um padrão da maneira como os dados foram gerados, portanto diversas imagens necessitavam de análise individual. A tabela 2 indica algumas informações sobre o dataset utilizado para o treinamento do modelo após o pré-processamento.

Após todo o tratamento do dataset houve uma diminuição expressiva do volume da base de dados, um pouco mais de 50% de redução, esse fator também gerou outros desafios durante a fase de treinamento, como a falta de elementos representativos. A imagem abaixo é um exemplo do resultado final sendo alimentado na rede.

Label	Amostras	Porcentagem
Psoríase	1756	54.9%
Dermatite	1071	45.1%
Total	2258	100%

Tabela 2 – Distribuição do dataset após o processo de curadoria.

Apesar da redução do volume de dados, o modelo alcançou um desempenho satisfatório. A Tabela 3 resume as métricas de avaliação obtidas no conjunto de teste. Observa-se que a acurácia de 93,78% e o F1-Score de 91,76% indicam um desempenho equilibrado, com a rede demonstrando robustez na distinção entre as classes.

Além da avaliação quantitativa, a interpretabilidade do modelo foi investigada por meio da técnica de *GradCAM*. A Figura 8 apresenta um mapa de calor que ilustra

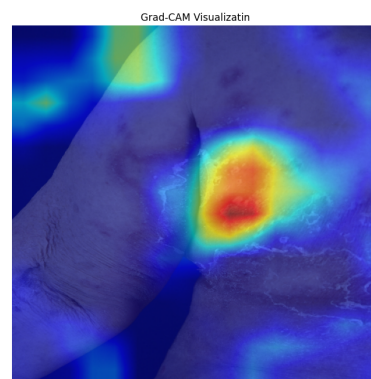
Métrica	Valor
Acurácia	93,78%
Precisão	92,11%
Revocação	91,42%
F1-Score	91,76%

Tabela 3 – Resultados das métricas de avaliação do modelo no conjunto de teste.

as regiões de uma imagem de psoríase nas quais o modelo concentrou sua atenção para realizar a classificação. A análise visual desta figura revela aspectos importantes sobre o comportamento de atenção do modelo. Observa-se que as regiões de maior ativação (destacadas em vermelho e amarelo) concentram-se especificamente nas bordas bem definidas da lesão psoriática, onde há uma clara demarcação entre a pele saudável e a área afetada.



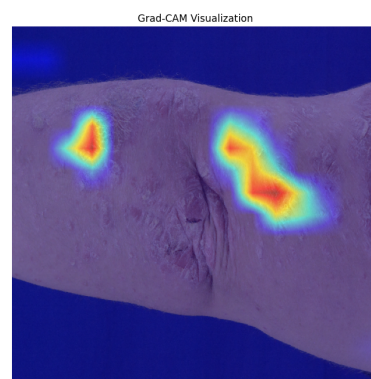
(a) Imagem original - Psoríase



(b) GradCAM - Psoríase



(c) Imagem original - Psoríase



(d) GradCAM - Psoríase

Figura 8 – Resultados da análise GradCAM para classificação de psoríase e dermatite.

O modelo demonstra sensibilidade particular às variações cromáticas características da psoríase, focalizando sua atenção nas áreas de eritema (vermelhidão) e nas regiões de descamação esbranquiçada, que são sinais clínicos distintivos desta condição dermatológica. Adicionalmente, nota-se que a rede prioriza as características morfológicas da lesão, como

seu formato irregular e a textura heterogênea da superfície, aspectos fundamentais para o diagnóstico diferencial entre psoríase e outras dermatoses. Esta análise confirma que o modelo não se baseia em ruídos ou artefatos da imagem, mas sim aprende a identificar e priorizar as características visuais clinicamente relevantes, possibilitando sua aplicabilidade como ferramenta de apoio ao diagnóstico dermatológico.

Durante o treinamento e a avaliação foram utilizadas diferentes técnicas que possibilitassem a análise do desempenho do modelo, a figura 8 demonstra as áreas de interesse que o modelo analisou para a classificação da doença. A partir dessa figura, é possível compreender como o modelo foi capaz de relacionar o contexto da região vermelha (área lesionada) com a classificação da doença psoríase

5 Conclusão

Este trabalho apresentou o desenvolvimento de um sistema CAD para diagnóstico de psoríase vulgar baseado em *Visual Transformers*, utilizando dados de pacientes brasileiros fornecidos pela FMABC. O objetivo principal foi criar uma ferramenta de apoio ao diagnóstico capaz de distinguir entre psoríase vulgar e dermatite por meio de técnicas de visão computacional e processamento digital de imagens. Para atingir essa meta, foi implementado um *pipeline* completo que envolveu desde a coleta e curadoria dos dados até o treinamento e avaliação de um modelo baseado na arquitetura *Swin Transformers*. O modelo desenvolvido alcançou resultados promissores, com acurácia de 93,78%, precisão de 92,11%, revocação de 91,42% e F1-Score de 91,76%, demonstrando sua capacidade de classificação binária entre as duas condições dermatológicas estudadas.

Apesar dos resultados quantitativos satisfatórios, é fundamental destacar as limitações inerentes ao modelo, em grande parte, devido às características do dataset utilizado. A redução de mais de 50% da base de dados original, embora necessária para garantir a qualidade das amostras, resultou em um volume limitado de imagens para o treinamento.

Adicionalmente, as imagens foram capturadas em um ambiente controlado de laboratório universitário, o que as torna menos representativas de cenários clínicos reais, onde a variação de iluminação, qualidade da câmera e angulação é muito maior. Consequentemente, o modelo pode ter sua capacidade de generalização comprometida ao ser aplicado a imagens com características diferentes daquelas do conjunto de dados curado, como em um cenário de triagem em consultórios médicos. Tais limitações reforçam a

necessidade de futuras pesquisas que utilizem um conjunto de dados mais diversificado, de maior volume e que simulem as condições do mundo real, a fim de aprimorar a robustez e a aplicabilidade clínica do modelo.

Apesar das limitações inerentes ao conjunto de dados, os resultados obtidos nesta pesquisa demonstram o potencial significativo do desenvolvimento de um sistema CAD. A utilização de dados de pacientes brasileiros confere um valor estratégico ao modelo para a aplicabilidade em cenários clínicos nacionais. Além disso, este estudo serve como um ponto de partida para trabalhos futuros promissores. A evolução para um sistema capaz de realizar a segmentação automática da lesão ou a classificação da severidade da doença são direções de pesquisa que poderiam ampliar a funcionalidade e o impacto clínico da ferramenta.

Referências

- ARORA, R. et al. Automated skin lesion segmentation using attention-based deep convolutional neural network. *Biomedical Signal Processing and Control*, Elsevier BV, v. 65, p. 102358, mar. 2021. ISSN 1746-8094. Disponível em: <<http://dx.doi.org/10.1016/j.bspc.2020.102358>>.
- BARROSO, W. K. S. et al. Diretrizes brasileiras de hipertensão arterial – 2020. *Arquivos Brasileiros de Cardiologia*, Sociedade Brasileira de Cardiologia, v. 116, n. 3, p. 516–658, mar. 2021. ISSN 1678-4170. Disponível em: <<http://dx.doi.org/10.36660/abc.20201238>>.
- BRINKER, T. J. et al. Deep neural networks are superior to dermatologists in melanoma image classification. *European Journal of Cancer*, v. 119, p. 11–17, 2019. ISSN 0959-8049.
- DASH, M. et al. A cascaded deep convolution neural network based cadx system for psoriasis lesion segmentation and severity assessment. *Applied Soft Computing*, v. 91, p. 106240, 2020. ISSN 1568-4946.
- DOI, K. Computer-aided diagnosis in medical imaging: Historical review, current status and future potential. *Computerized Medical Imaging and Graphics*, v. 31, n. 4, p. 198–211, 2007. ISSN 0895-6111. Computer-aided Diagnosis (CAD) and Image-guided Decision Support. Disponível em: <<https://www.sciencedirect.com/science/article/pii/S0895611107000262>>.
- DOSOVITSKIY, A. et al. An image is worth 16x16 words: Transformers for image recognition at scale. arXiv, 2020. Disponível em: <<https://arxiv.org/abs/2010.11929>>.
- GONZALEZ, R.; WOODS, R. *Digital Image Processing*. [S.l.]: Prentice Hall, 2008. ISBN 9780131687288.
- GRIFFITHS, C. E.; BARKER, J. N. Pathogenesis and clinical features of psoriasis. *The Lancet*, Elsevier BV, v. 370, n. 9583, p. 263–271, jul. 2007. ISSN 0140-6736. Disponível em: <[http://dx.doi.org/10.1016/S0140-6736\(07\)61128-3](http://dx.doi.org/10.1016/S0140-6736(07)61128-3)>.

HAMEED, N. et al. An intelligent computer-aided scheme for classifying multiple skin lesions. *Computers*, MDPI AG, v. 8, n. 3, p. 62, ago. 2019. ISSN 2073-431X. Disponível em: <<http://dx.doi.org/10.3390/computers8030062>>.

HAYKIN, S. *Neural Networks and Learning Machines*. [S.l.]: Pearson, 2009. (Pearson International Edition). ISBN 9780131293762.

MARTINIANO, A. et al. Utilizando uma rede neural artificial para aproximação da função de evolução do sistema de lorentz. *Revista Produção e Desenvolvimento*, v. 2, p. 26–38, 04 2016.

MILLER, I. M. et al. Meta-analysis of psoriasis, cardiovascular disease, and associated risk factors. *Journal of the American Academy of Dermatology*, Elsevier BV, v. 69, n. 6, p. 1014–1024, dez. 2013. ISSN 0190-9622. Disponível em: <<http://dx.doi.org/10.1016/j.jaad.2013.06.053>>.

RUSSELL, S.; NORVIG, P. *Artificial Intelligence: A Modern Approach*. 3rd. ed. USA: Prentice Hall Press, 2009. ISBN 0136042597.

VASWANI, A. et al. Attention is all you need. arXiv, 2017. Disponível em: <<https://arxiv.org/abs/1706.03762>>.