

Câncer de Pele

MNIST: HAM10000



Trabalho da disciplina de Tópicos especiais em IA

Estudante: Anderson Oliveira, Bárbara Talita, Jeovane Santos, Quele Andrade, Ricardo Teixeira

Orientador: Dr. Tiago Pagano

Curso: Bacharelado em Engenharia de Computação

16 de dezembro de 2024

Sumário

- 1 **Resumo**
- 2 **Introdução**
 - Contextualização
 - Justificativa
 - Objetivo
- 3 **Metodologia**
 - Dataset
 - Pré-processamento
 - Arquitetura
- Desbalanceamento das classes no dataset
- 4 **Resultados esperados**
 - Resultados esperados
- 5 **Resultados Parciais**
 - 1ª arquitetura utilizada
 - 2ª arquitetura utilizada
 - 3ª arquitetura utilizada
- 6 **Considerações finais**
 - Considerações finais

Resumo

O dataset Skin Cancer MNIST: HAM10000 contém 10.015 imagens de lesões cutâneas divididas em sete classes, incluindo melanoma e outras doenças da pele, com diagnósticos validados por especialistas.

Palavras-Chave: Câncer de pele; Melanoma; Machine Learning; Inteligência Artificial; Diagnóstico médico.

Contextualização

- O câncer de pele é uma das doenças mais comuns no mundo, com altos índices de incidência e desafios no diagnóstico precoce, especialmente em regiões com acesso limitado a dermatologistas.
- Algoritmos de aprendizado profundo podem identificar padrões morfológicos em imagens histológicas para classificar subtipos de câncer, como os baseados nos receptores ER, PR e Her2 [1].
- Em áreas com classes desbalanceadas, como o câncer de pele, é essencial usar métodos que aprimorem a precisão dos modelos de diagnóstico, minimizando falsos positivos e negativos [2].

Justificativa

- O câncer de pele apresenta alta prevalência e pode levar a desfechos graves quando não diagnosticado precocemente, especialmente em casos como o melanoma.
- A inteligência artificial tem o potencial de melhorar a precisão e a acessibilidade do diagnóstico dermatológico, mesmo em locais com recursos limitados.

Objetivo

Desenvolver um modelo de inteligência artificial capaz de identificar e classificar diferentes tipos de câncer de pele e outras doenças dermatológicas a partir de imagens.

Metodologia: Dataset

- Contém 10.015 imagens de alta qualidade de lesões cutâneas.
- Abrange sete classes distintas: Melanocytic nevi, Melanoma, Benign keratosis-like lesions, Basal cell carcinoma, Actinic keratoses, Vascular lesions, e Dermatofibroma.
- Dados fornecidos em formato JPEG com metadados estruturados em arquivos CSV.
- Inclui imagens de pacientes de diversas origens demográficas.

Metodologia: Pré-processamento

- As imagens são organizadas por tipo de câncer, com cada classe de lesão cutânea em pastas específicas.
- O conjunto de dados é dividido em três partes: validação, treino e teste.

Metodologia: Arquitetura

- Utilizamos a arquitetura InceptionV3 por sua capacidade de capturar padrões complexos em imagens.
- Aplicamos Dropout para prevenir overfitting e melhorar a generalização do modelo.
- Realizamos fine-tuning.
- A avaliação do modelo foi feita usando AUC-PRC, pois as classes estavam desbalanceadas.

Metodologia: Arquitetura

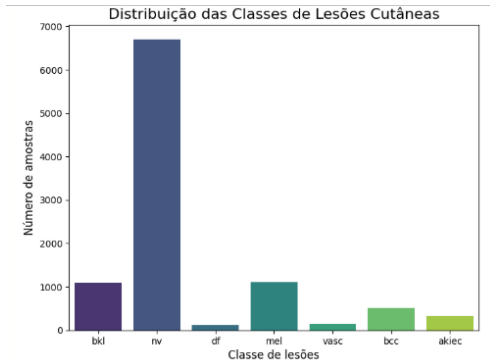


Figura: Desbalanceamento das classes

Resultados esperados

- Espera-se que o modelo seja capaz de identificar e classificar corretamente diferentes tipos de câncer de pele e outras doenças dermatológicas com alta precisão.
- O uso de AUC-PRC como métrica de avaliação deve garantir que o modelo consiga lidar com dados desbalanceados, proporcionando um bom equilíbrio entre precisão e recall.

Resultados Parciais: 1ª arquitetura utilizada

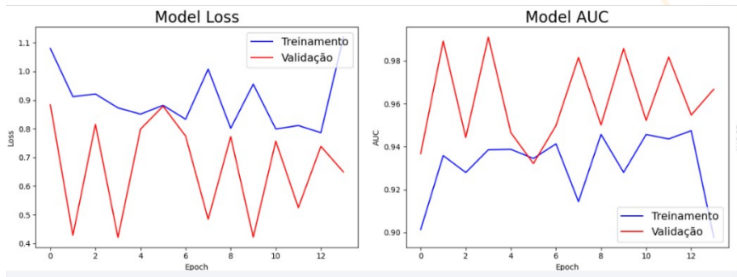


Figura: Primeira arquitetura

Resultados Parciais: 2ª arquitetura utilizada

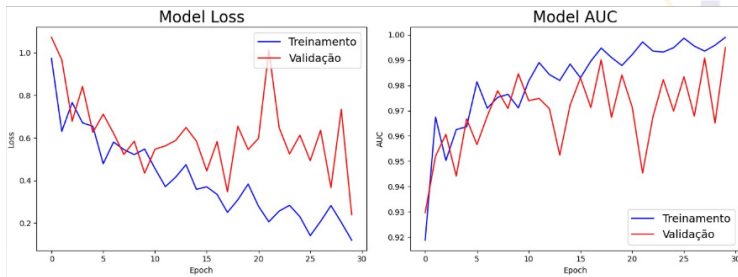


Figura: Segunda arquitetura

Resultados Parciais: 3ª arquitetura utilizada

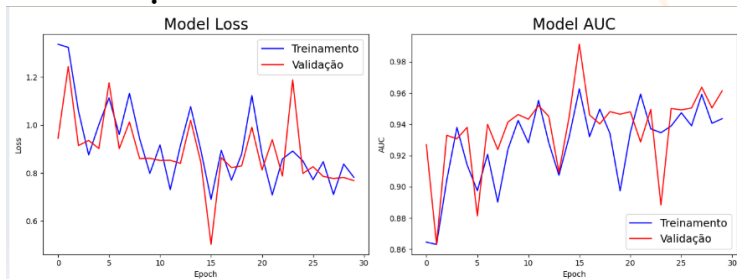


Figura: Terceira arquitetura

Considerações finais

- Base de dados robusta e diversificada para desenvolvimento de modelos.
- Metodologia mostrou-se ideal para classificação de lesões cutâneas.

Referências I

- [1] Rishi R. Rawat et al. “Deep learned tissue “fingerprints” classify breast cancers by ER/PR/Her2 status from H&E images”. Em: **Scientific Reports** 10 (2020). DOI: 10 . 1038 / s41598 - 020 - 63774 - 2. URL: <https://www.nature.com/articles/s41598-020-63774-2>.
- [2] Shahzad Ali Khan e Zeeshan Ali Rana. “Evaluating performance of software defect prediction models using area under precision-Recall curve (AUC-PR)”. Em: **2019 2nd International Conference on Advancements in Computational Sciences (ICACS)**. IEEE, 2019. DOI: 10 . 1109 / ICACS . 2019 . 00054. URL: <https://ieeexplore.ieee.org/document/8884546>.

Obrigado(a) pela Atenção!

Contato:

`anderson.oliveira@aluno.ufrb.edu.br`

`barbarans@aluno.ufrb.edu.br`

`jeovanessantos@aluno.ufrb.edu.br`

`queleandrade@aluno.ufrb.edu.br`

`ricardoteixeirasantos@aluno.ufrb.edu.br`