Classificação de Imagens PET de Corpo Inteiro com FDG-18 para Diagnóstico de Câncer

Celso Luiz Silva Soares Filho[1]

[1]Programa de Pós-Graduação em Engenharia Elétrica – Universidade Federal do Maranhão (UFMA)
Av. dos Portugueses, 1966 - Vila Bacanga, São Luís - MA, 65080-805

celso.soares@discente.ufma.br

Resumo. *A tomografia por emissão de pósitrons (PET) faz parte da área de medicina nuclear e tem se mostrado extremamente útil para a detecção e o acompanhamento de patologias. Com as imagens geradas a partir dos exames PET, em conjunto com o radiofármaco 18F-fluorodesoxiglicose (FDG-18), é possível visualizar áreas com maior atividade metabólica no corpo, como o cérebro, a bexiga urinária, inflamações e células cancerosas. Este trabalho tem como objetivo classificar as imagens de exames PET entre pacientes saudáveis e pacientes com câncer (como linfoma, melanoma e câncer de pulmão), comparando os resultados entre diferentes técnicas, como XGBoost, Redes Neurais Siamesas e Vision Transformer. Resultados promissores foram alcançados na classificação dessas imagens.*

1. Introdução

O câncer é uma doença muito preocupante, que pode se manifestar de diversas formas diferentes, como linfoma, melanoma e câncer de pulmão, cada uma com suas particularidades para o tratamento e diagnóstico. No Brasil, os cânceres são frequentemente diagnosticados, representando um desafio significativo para a saúde pública [Corrêa 2023].

O diagnóstico e o tratamento do câncer enfrentam desafios significativos devido à grande quantidade de fatores associados a cada tipo da doença, além da complexidade envolvida no diagnóstico precoce [Corrêa 2023]. Nesse contexto, a Tomografia por Emissão de Pósitrons (PET) desempenha um papel crucial no diagnóstico e no monitoramento do câncer, pois essa tecnologia permite visualizar com precisão a atividade metabólica dos tecidos, auxiliando na identificação de áreas de crescimento anormal e no acompanhamento do tratamento [da Silva et al. 2024].

O exame de PET é uma ferramenta de imagem amplamente utilizada no diagnóstico de diversos tipos de câncer [Savoie et al. 2022]. Ele é realizado com a aplicação de um radiofármaco, como o 18F-fluorodesoxiglicose (FDG-18), marcado com flúor-18, que possui um comportamento análogo ao da glicose, concentrando-se em áreas com maior atividade metabólica no corpo, como o cérebro, a bexiga urinária, os rins e tumores [Duclos et al. 2021]. Com a concentração do radiofármaco nessas áreas, o decaimento do composto radioativo (flúor-18) é detectado pelo aparelho PET, gerando imagens detalhadas que permitem localizar e avaliar tumores e metástases, bem como acompanhar a resposta ao tratamento realizado no paciente [Savoie et al. 2022].

As imagens geradas a partir do PET podem ser utilizadas em tarefas de Visão Computacional, como classificação e segmentação, com o auxílio de diversas técnicas,

tais como Redes Neurais Convolucionais (CNN), XGBoost, Redes Neurais Siamesas, Transformers, entre outras, para apoiar os médicos no diagnóstico dos pacientes [Jiang et al. 2024]. Assim, este trabalho tem como objetivo aplicar as técnicas de XGBoost, Redes Neurais Siamesas e Vision Transformer para a classificação de exames PET, comparando os resultados obtidos na detecção de câncer, como linfoma, melanoma ou câncer de pulmão, e identificando se o paciente está saudável ou não.

Este trabalho está organizado da seguinte forma: a Seção 1 apresenta a motivação e objetivo. A Seção 2 detalha a base de imagens e o método utilizado. A Seção 3 apresenta os resultados da classificação. A Seção 4 apresenta as conclusões do trabalho.

2. Materiais e Método

2.1. Base de Imagens

A base de imagens utilizada neste trabalho foi adquirida a partir do AutoPet Challenge III [III 2024, Gatidis et al. 2022]. Esse desafio consiste em uma competição voltada para a segmentação de lesões relacionadas a câncer de pulmão, linfoma, melanoma e pacientes saudáveis, utilizando exames 3D de tomografia computadorizada (CT), PET e anotações de especialistas contendo as máscaras das lesões. A base de dados inclui exames realizados com o radiofármaco 18F-fluorodesoxiglicose (FDG-18), totalizando 1014 exames de 900 pacientes, além de exames realizados com o Prostate-specific Membrane Antigen (PSMA), somando 597 exames de 378 pacientes.

As imagens dos exames PET estão em nível de SUV (Standardized Uptake Value), uma medida que avalia a atividade metabólica do corpo, com valores mais elevados nas regiões com maior atividade metabólica. As dimensões dos volumes variam de acordo com o tipo de exame, uma vez que alguns são realizados do topo da cabeça até a metade superior das coxas dos pacientes, enquanto outros abrangem o corpo inteiro [III 2024, Gatidis et al. 2022].

Para o presente trabalho, foram utilizadas exclusivamente as imagens PET obtidas com o radiofármaco FDG-18, buscando uma menor complexidade inicial na exploração dos dados. Além disso, as fatias individuais do volume dos exames não foram usadas; em vez disso, foi adotada uma representação 2D baseada nos valores de MIP (Maximum Intensity Projection). As imagens MIP representam o volume completo do exame, sendo geradas a partir dos valores mais altos de SUV (Standardized Uptake Value) ao longo de todo o volume [Kawakami et al. 2020].

As imagens passaram por um processo de redimensionamento, onde todas foram ajustadas para 400 pixels de altura e 400 pixels de largura. Caso a altura original fosse menor que o valor determinado, pixels pretos foram adicionados nas partes superior e inferior, preservando as características das imagens e evitando distorções. Para imagens com altura maior que a especificada, foi realizado um corte na parte inferior, mantendo-se a região superior, uma vez que a remoção das pernas não prejudica o desempenho dos experimentos, considerando que essa região não é determinante para melhores resultados.

Para padronizar a intensidade dos pixels nas imagens MIP, foi adotado um valor de SUV igual a 15 como referência para o valor máximo, assegurando consistência nas representações MIP. Essas modificações nas imagens foram realizadas com base em experimentos conduzidos e nas recomendações de classificação apresentadas em

[Heiliger et al. 2022]. A Figura 1 apresenta um exemplo de uma imagem MIP utilizada neste trabalho.

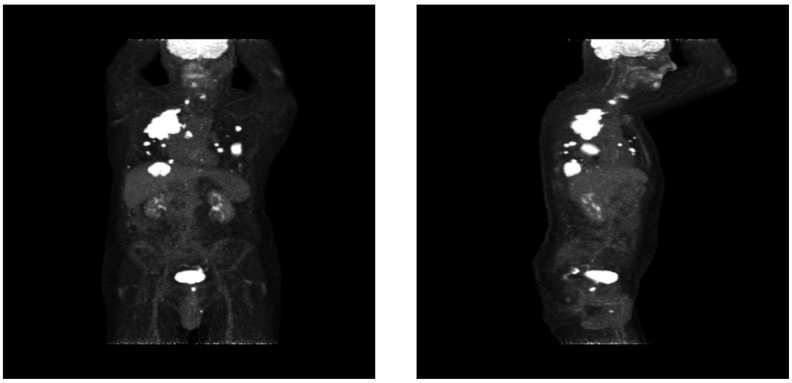

Figura 1. Representações MIP dos Cortes Coronal e Sagital de uma paciente com câncer de pulmão. Fonte: Elaborado pelo Autor.

2.2. XGBoosting

O XGBoost (Extreme Gradient Boosting) é uma técnica de aprendizado supervisionado baseada no algoritmo de boosting por gradiente [Chen and Guestrin 2016]. Ele é amplamente reconhecido por sua alta eficiência, flexibilidade e portabilidade, sendo muito utilizado na construção de modelos preditivos para problemas de regressão e classificação [Chen and Guestrin 2016].

O XGBoost utiliza o conceito de ensemble learning, combinando múltiplos modelos fracos, como árvores de decisão, para formar um modelo forte. O boosting por gradiente é um método iterativo que ajusta sucessivamente as árvores de decisão para corrigir os erros cometidos pelos modelos anteriores. Cada nova árvore é construída com base no gradiente do erro, permitindo ao algoritmo concentrar-se nas previsões mais difíceis e, assim, melhorar progressivamente a precisão global do modelo. Essa abordagem iterativa e acumulativa é o que torna o XGBoost uma ferramenta poderosa e eficaz na solução de problemas complexos [Chen and Guestrin 2016].

Neste trabalho, o XGBoost é utilizado para a classificação das features extraídas das imagens. Inicialmente, uma CNN é empregada para a extração das características mais relevantes das imagens, e, em seguida, essas características são classificadas utilizando o XGBoost. O backbone da ResNet-18 foi adotado para a extração de características, com base na metodologia de classificação apresentada por Heiliger et al. (2022).

2.3. Redes Neurais Siamesas

Outro método de classificação empregado neste trabalho foi a Rede Neural Siamesa (Siamese Neural Network, SNN). Essas redes são projetadas para aprender representações que permitam a comparação entre pares de entradas. Elas consistem em duas (ou mais) sub-redes idênticas que compartilham os mesmos pesos e parâmetros, garantindo que ambas processem as entradas de maneira simétrica e aprendam representações consistentes [Bromley et al. 1993].

As Redes Neurais Siamesas são especialmente eficazes em tarefas de verificação de similaridade, cujo objetivo é determinar se duas entradas pertencem à mesma classe ou estão relacionadas de alguma forma. Em vez de classificar diretamente os dados, essas redes aprendem a medir a similaridade entre pares de entradas, utilizando uma função de distância, como a distância euclidiana ou a função de contraste [Bromley et al. 1993].

Neste trabalho, a implementação da Rede Neural Siamesa utilizou o backbone ResNet-18 para extrair características das imagens PET. Para o treinamento, foi adotada a Focal Loss como função de perda, devido à sua capacidade de lidar com dados desbalanceados, enfatizando os exemplos mais difíceis. O otimizador escolhido foi o AdamW, que incorpora regularização baseada em decaimento de peso, contribuindo para prevenir overfitting. O modelo foi treinado com um batch size de 32, uma taxa de aprendizado inicial de 1e-4, e as representações aprendidas foram ajustadas para medir a similaridade entre pares de imagens PET.

2.4. ResNet-18 como Backbone

A arquitetura da ResNet-18 foi utilizada neste trabalho como backbone para os experimentos realizados com as técnicas de XGBoost e Redes Neurais Siamesas. O principal objetivo de sua utilização é a extração de features (características), que são empregadas posteriormente para a classificação em ambos os casos.

A arquitetura, composta por camadas residuais, foi proposta por He et al. (2016) e se destaca por facilitar o treinamento de redes profundas, resolvendo problemas como o desaparecimento de gradientes e a degradação do desempenho em redes muito profundas. A Figura 2 ilustra o funcionamento do bloco residual, elemento fundamental dessa arquitetura.

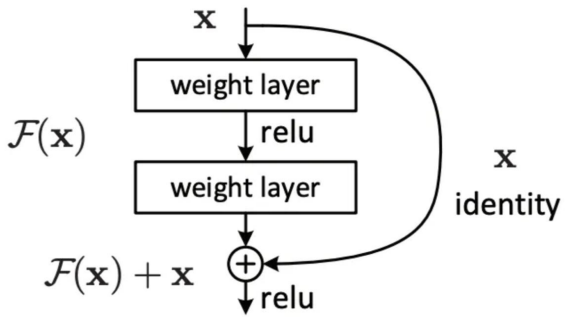

Figura 2. Bloco Residual [He et al. 2016].

Os blocos residuais, ilustrados na Figura 2, têm como função principal aprender incrementos F(x) em vez de um mapeamento completo H(x) . Para isso, utilizam conexões de atalho que preservam a entrada original x e a somam ao resultado processado. Essa abordagem proporciona maior estabilidade durante o treinamento, facilita a propagação eficiente de gradientes e assegura que informações importantes não sejam perdidas [He et al. 2016].

2.5. Vision Transformer (ViT)

O Vision Transformer (ViT), introduzido por Dosovitskiy et al. (2020), representa uma abordagem inovadora no campo da visão computacional. Diferentemente das Redes Neu-

rais Convolucionais (CNNs), amplamente utilizadas em tarefas de visão computacional, o ViT adapta o modelo Transformer, originalmente desenvolvido para Processamento de Linguagem Natural (PLN), para lidar com imagens [Dosovitskiy 2020]. Essa adaptação trouxe o poderoso mecanismo de atenção, permitindo capturar relações globais nas imagens.

A arquitetura do ViT inicia dividindo as imagens em pequenos blocos, conhecidos como patches, geralmente de dimensões 16×16 pixels. Cada patch é linearizado (transformado em um vetor) e enriquecido com embeddings de posição, para preservar a ordem espacial. Esses vetores, que funcionam de maneira semelhante aos tokens no NLP, são processados por camadas de Transformer, compostas por mecanismos de atenção multi-cabeças (multi-head self-attention) e redes feedforward. Essa abordagem elimina a necessidade de convoluções, aprendendo o contexto global diretamente através dos mecanismos de atenção [Dosovitskiy 2020]. A Figura 3 ilustra a arquitetura do ViT, destacando o fluxo de processamento desde a entrada dos patches até a saída final do modelo.

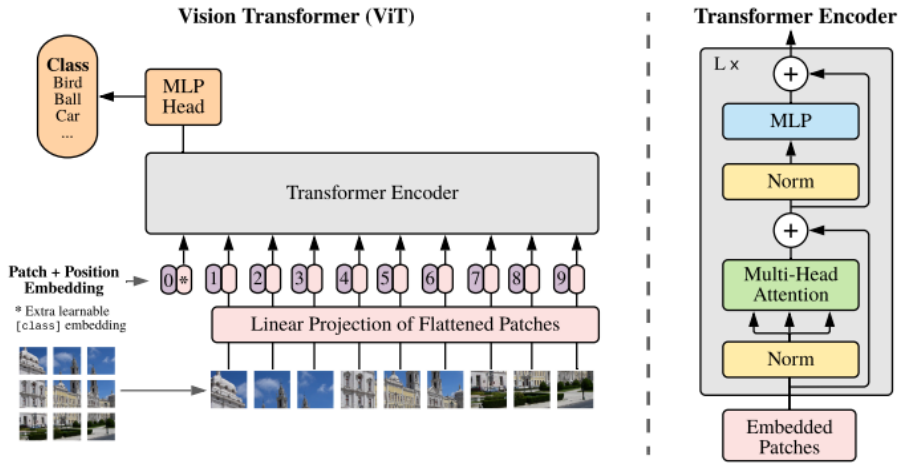

Figura 3. Arquitetura do ViT. Fonte [Dosovitskiy 2020].

A arquitetura do Vision Transformer (ViT), ilustrada na Figura 2, inicia dividindo a imagem em pequenos blocos chamados de patches, que são linearizados em vetores e enriquecidos com informações de posição, denominadas embeddings posicionais. Esses vetores são então processados por camadas Transformer, onde o mecanismo de atenção identifica as relações globais entre os patches. Ao final, um token específico é extraído para representar a classificação da imagem.

3. Resultados e Discussão

Para o treinamento das técnicas, foi utilizado o método de cross-validation, que divide a base de imagens em múltiplos folds, onde um fold é usado para validação e os demais para treinamento. Esse procedimento permite que o modelo seja treinado e avaliado utilizando todas as imagens disponíveis. Neste trabalho, o número de folds foi definido como 5, e o desempenho dos modelos foi avaliado por meio das métricas de F1-Score, Precisão, Recall e Acurácia.

A escolha do backbone foi com base em diversos experimentos, além de que Heiliger et al. 2022 utilizou a ResNet-18 com a intenção de diminuir ao máximo a quanti-

dade de Falsos Negativos. Os resultados médios, juntamente com os respectivos desvios padrão, para o corte coronal estão apresentados na Tabela 1. Já a Tabela 2 exibe as mesmas métricas de avaliação, mas para o corte sagital. Essas tabelas resumem a performance dos modelos, fornecendo uma análise comparativa das métricas nos diferentes planos de corte, permitindo avaliar a eficácia das técnicas propostas.

Tabela 1. Avaliação para os Corte Coronal

Método	Acurácia	F1-Score	Precisão	Recall
XGBoosting	0.7327 ± 0.0198	0.7322 ± 0.0196	0.7340 ± 0.0206	0.7327 ± 0.0198
SNN	0.5084 ± 0.0185	0.5826 ± 0.0130	0.5066 ± 0.0142	0.6864 ± 0.0318
ViT	**0.8294 ± 0.0194**	**0.8222 ± 0.0243**	**0.8474 ± 0.0345**	**0.8026 ± 0.0575**

Tabela 2. Avaliação para os Corte Sagital

Método	Acurácia	F1-Score	Precisão	Recall
XGBoosting	0.7554 ± 0.0291	0.7538 ± 0.0286	0.7618 ± 0.0339	0.7554 ± 0.0291
SNN	0.4921 ± 0.0198	0.5575 ± 0.0268	0.4946 ± 0.0166	0.6429 ± 0.0698
ViT	**0.7998 ± 0.0148**	**0.7895 ± 0.0228**	**0.8228 ± 0.0385**	**0.7646 ± 0.0633**

Os resultados obtidos demonstram um bom desempenho do Vision Transformer (ViT) em comparação com as outras técnicas utilizadas para a classificação de imagens PET com FDG-18 no conjunto de dados do AutoPet Challenge III. Para o corte coronal, os resultados do ViT destacam-se, especialmente pela alta Precisão, evidenciando a capacidade do modelo em minimizar falsos positivos e identificar corretamente as classes nesse plano.

Por outro lado, enquanto o XGBoost apresentou um desempenho consistente e próximo ao ViT, a Rede Neural Siamesa (SNN) mostrou resultados significativamente inferiores em ambos os cortes analisados. Apesar dos bons resultados gerais, todos os modelos apresentaram espaço para melhorias, especialmente no Recall, que foi o mais impactado, sugerindo maior dificuldade dos modelos em minimizar falsos negativos.

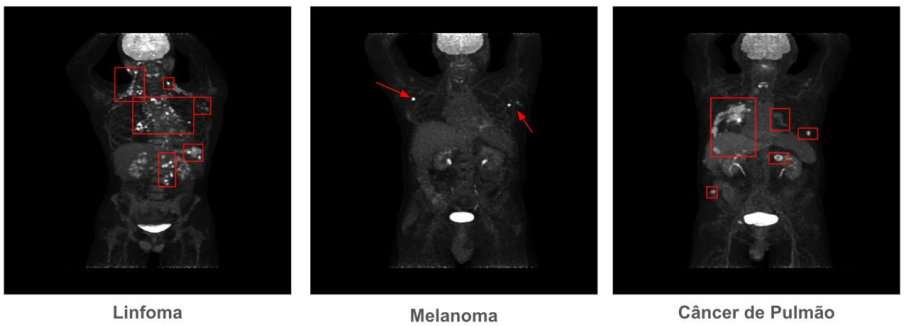

Figura 4. Casos onde todos acertaram. Fonte: Elaborado pelo Autor

Na Figura 4, é possível observar exemplos em que os três métodos classificaram corretamente, como verdadeiro positivo. Isso pode ocorrer devido ao fato de que o linfoma aparece com destaque na imagem, apresentando uma intensidade que se sobressai em relação às outras partes do corpo. No caso do exemplo de melanoma, embora a

área afetada pela patologia seja pequena, o melanoma é um tipo de câncer muito agressivo, com alta atividade metabólica nas regiões acometidas. Por outro lado, o câncer de pulmão pode ser diagnosticado a partir de baixos níveis de SUV, que frequentemente são consideravelmente menores do que os de outras áreas que sempre apresentam ativação, como a bexiga urinária e o cérebro.

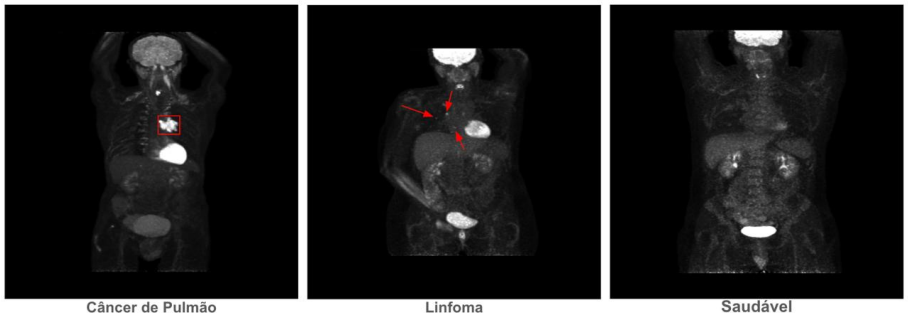

Figura 5. Casos onde apenas o ViT acertou. Fonte: Elaborado pelo Autor

Na Figura 5, é possível observar casos em que apenas o ViT classificou as imagens corretamente, sendo verdadeiro positivo no exemplo de câncer de pulmão e linfoma, e verdadeiro negativo na imagem do paciente saudável. Isso evidencia a capacidade do ViT de classificar corretamente imagens em que a patologia é muito discreta ou está localizada em regiões próximas a grandes fontes de falsos positivos, como o coração.

Figura 6. Caso onde apenas o ViT classificou como Falso Positivo. Fonte: Elaborado pelo Autor

No caso da Figura 6, apenas o ViT apresentou erro. Isso pode ter ocorrido porque o corpo inteiro apresenta uma grande região de ativação, especialmente próxima à bexiga. Diferentemente disso, a Figura 7 apresenta exemplos mais preocupantes, nos quais apenas o ViT classificou como falso negativo.

Na Figura 8, é possível observar imagens em que todos os métodos apresentaram erros. Em todos os casos, a patologia é muito discreta e se confunde significativamente com outras áreas do corpo, dificultando a classificação automática.

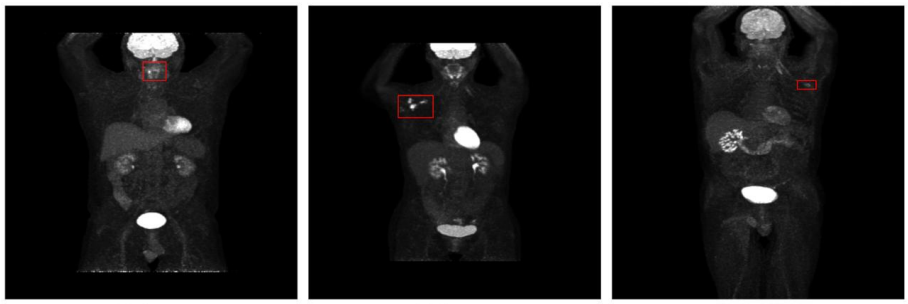

Figura 7. Caso onde apenas o ViT classificou como Falso Negativo. Fonte: Elaborado pelo Autor

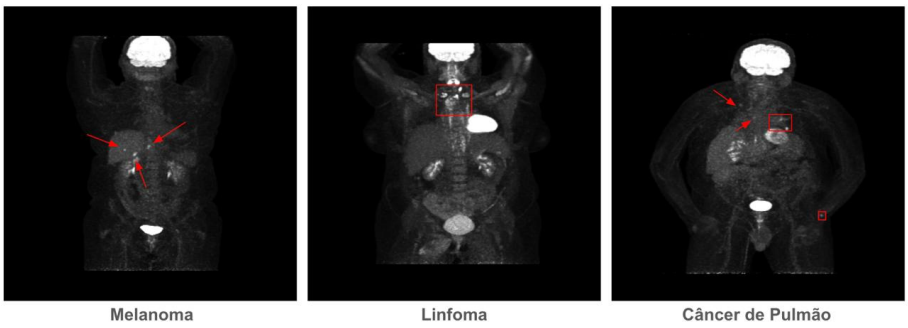

Figura 8. Caso onde apenas o ViT classificou como Falso Negativo. Fonte: Elaborado pelo Autor

De forma geral, os resultados demonstram a eficácia do modelo ViT, com métricas robustas e desvios padrão relativamente baixos, indicando consistência no desempenho. No entanto, os dados sugerem que ajustes no treinamento, como a inclusão de novas imagens ou o uso de técnicas de data augmentation, podem contribuir para aprimorar ainda mais o desempenho dos modelos, especialmente na detecção de classes menos representadas.

Em comparação com a literatura e trabalhos que utilizaram o mesmo conjunto de dados, é possível observar na Tabela 3 que o melhor método deste estudo, que utilizou o ViT, também se destaca em relação aos outros trabalhos. O estudo de Pang et al. (2024) buscou detectar patologias combinando abordagens 2D e 3D, enquanto Heiliger et al. (2022) utilizou uma combinação entre ResNet-18 e ResNet-50 nos planos coronal e sagital, com o objetivo de minimizar ao máximo a quantidade de falsos negativos.

Tabela 3. Comparação com a literatura

Método	Acurácia	F1-Score	Precisão	Recall
Pang et al. (2024)	0.78 ± 0.28	-	0.84 ± 0.19	**0.98 ± 0.06**
Heiliger et al. (2022)	0.743	-	-	-
Método Proposto Coronal / Sagital	**0.8294 ± 0.0194/ 0.7998 ± 0.0148**	**0.8222 ± 0.0243/ 0.7895 ± 0.0228**	**0.8474 ± 0.0345/ 0.8228 ± 0.0385**	0.8026 ± 0.0575/ 0.7646 ± 0.0633

4. Conclusão

Este projeto de mestrado utilizou o XGBoost, as Redes Neurais Siamesas (SNN) e o Vision Transformer (ViT) para a classificação de imagens PET com FDG-18 do conjunto de dados AutoPet Challenge III. Os resultados alcançados foram promissores para os cortes Coronal e Sagital, mas ainda apresentam espaço para melhorias, como a inclusão de todas as imagens PET disponíveis no conjunto de dados e a aplicação de técnicas que possam aprimorar as métricas de avaliação, como o data augmentation.

Para trabalhos futuros, propõe-se a inclusão de imagens PET obtidas com o radiofármaco PSMA, o que pode melhorar as métricas devido à maior diversidade de dados. Além disso, o uso de técnicas de data augmentation tem o potencial de aumentar significativamente a variedade de imagens disponíveis para o treinamento, ampliando a robustez dos modelos. Outra abordagem promissora é a combinação (ensemble) de redes treinadas para os planos Coronal e Sagital, permitindo uma solução mais abrangente e robusta.

Adicionalmente, recomenda-se a realização de novos experimentos com redes neurais baseadas em Transformer, como o Swin Transformer, que podem capturar características mais complexas. Por fim, sugere-se explorar backbones mais avançados para a extração de características, potencializando o desempenho do XGBoost na classificação. Essas direções futuras prometem contribuir para avanços significativos na classificação de imagens PET e no diagnóstico auxiliado por visão computacional.

Referências

Bromley, J., Guyon, I., LeCun, Y., Säckinger, E., and Shah, R. (1993). Signature verification using a"siamese"time delay neural network. *Advances in neural information processing systems*, 6.

Chen, T. and Guestrin, C. (2016). Xgboost: A scalable tree boosting system. In *Proceedings of the 22nd acm sigkdd international conference on knowledge discovery and data mining*, pages 785–794.

Corrêa, P. B. (2023). Perfil das pessoas que realizaram ou estão em tratamento oncológico em relação à adesão ao aconselhamento nutricional, incluindo uso de polifenóis.

da Silva, C. F. M., Zabot, J. B., and de Macena Alves, A. (2024). Utilização do pet-ct em recorrências do câncer de mama: revisão sistemática. *TCC-Biomedicina*.

Dosovitskiy, A. (2020). An image is worth 16x16 words: Transformers for image recognition at scale. *arXiv preprint arXiv:2010.11929*.

Duclos, V., Iep, A., Gomez, L., Goldfarb, L., and Besson, F. L. (2021). Pet molecular imaging: a holistic review of current practice and emerging perspectives for diagnosis, therapeutic evaluation and prognosis in clinical oncology. *International journal of molecular sciences*, 22(8):4159.

Gatidis, S., Hepp, T., Früh, M., La Fougère, C., Nikolaou, K., Pfannenberg, C., Schölkopf, B., Küstner, T., Cyran, C., and Rubin, D. (2022). A whole-body fdg-pet/ct dataset with manually annotated tumor lesions. *Scientific Data*, 9(1):601.

He, K., Zhang, X., Ren, S., and Sun, J. (2016). Deep residual learning for image recognition. In *Proceedings of the IEEE conference on computer vision and pattern recognition*, pages 770–778.

Heiliger, L. et al. (2022). Autopet challenge: combining nn-unet with swin unetr augmented by maximum intensity projection classifier (2022). *DOI: https://doi.org/10.48550/ARXIV*, 2209.

III, G. C. A. (2024). Automated lesion segmentation in whole-body pet/ct - multitracer multicenter generalization. `https://autopet-iii.grand-challenge.org/`. Acessado em 19 de dezembro de 2024.

Jiang, X., Wang, S., and Zhang, Y. (2024). Vision transformer promotes cancer diagnosis: A comprehensive review. *Expert Systems with Applications*, 252:124113.

Kawakami, M., Hirata, K., Furuya, S., Kobayashi, K., Sugimori, H., Magota, K., and Katoh, C. (2020). Development of combination methods for detecting malignant uptakes based on physiological uptake detection using object detection with pet-ct mip images. *Frontiers in medicine*, 7:616746.

Savoie, P., Murez, T., Neuville, P., Van Hove, A., Rocher, L., Fléchon, A., Camparo, P., Ferretti, L., Branger, N., and Rouprêt, M. (2022). French afu cancer committee guidelines update 2022–2024: Adrenal tumor–assessment of an adrenal incidetaloma and oncological management. *Progrès en Urologie*, 32(15):1040–1065.