

Classificação da Doença Hepática Esteatótica Associada a Disfunção Metabólica em Imagens Térmicas Utilizando Séries Temporais

Wenderson Arthur Dutra Oliveira¹, Francisco Roberto Cantanhede Brito¹,
Daniel Moreira Pinto¹, Aristófaes Corrêa Silva¹,
Henrique Manoel de Araujo Martins Filho^{2,3}

¹Núcleo de Computação Aplicada - Universidade Federal do Maranhão (UFMA)
Av. dos Portugueses, 1966 - Vila Bacanga, São Luís - MA, 65080-805

²Universidade Anhembi Morumbi - UAM
Rua Casa do Ator, 294 - Vila Olímpia, São Paulo - SP, Brasil, 04546-001

³Alchimia
Rua Mar. Castelo Branco 31, Olho D'água, São Luís, MA, Brasil, 65065-090

{wenderson.arthur, francisco.rcb}@discente.ufma.br,
{daniel.pinto, ari}@nca.ufma.br, {henriquealchimia}@hotmail.com

Abstract. *Metabolic dysfunction-associated steatotic liver disease affects approximately 30% of the global population. Its clinical course may progress to severe conditions such as cirrhosis and hepatocellular carcinoma. In this context, thermography has emerged as a non-invasive, yet challenging, method for screening for the disease. This study compared temporal, spatial, and spatio-temporal integration approaches for liver disease classification using thermal image time series. The experiments achieved a precision of 66% and sensitivity of 73%, demonstrating that preserving temporal evolution along with spatial patterns results in better performance.*

Resumo. *A doença hepática esteatótica associada à disfunção metabólica afeta aproximadamente 30% da população global. Seu curso clínico pode progredir para condições severas, como cirrose e carcinoma hepatocelular. Nesse contexto, a termografia tem se destacado como um método não invasivo, porém desafiador, para o rastreamento da doença. Este estudo comparou abordagens temporais, espaciais e integração espaço-temporal para a classificação da doença hepática utilizando séries temporais em imagens térmicas. Os experimentos alcançaram resultados de 66% de precisão e 73% de sensibilidade, demonstrando que a preservação da evolução temporal junto com os padrões espaciais resulta em um melhor desempenho.*

1. Introdução

A doença hepática esteatótica associada à disfunção metabólica (MASLD, do inglês *Metabolic dysfunction-Associated Steatotic Liver Disease*), caracteriza-se pela presença de esteatose hepática em indivíduos com pelo menos um fator de risco cardiometabólico, na ausência de consumo alcoólico prejudicial [EASL et al. 2024]. A doença

abrange um espectro que varia da esteatose simples até formas mais graves, como fibrose, cirrose e carcinoma hepatocelular [EASL et al. 2024, Friedman et al. 2018]. Devido ao seu curso inicial silencioso e à alta prevalência global, estimada em 30% da população adulta, a doença representa atualmente um importante problema de saúde pública [Amini-Salehi et al. 2024]. Tradicionalmente, a biópsia hepática é o método mais eficaz para diagnóstico da doença. No entanto, ela possui limitações, sendo um procedimento invasivo, sujeitos a riscos clínicos e vieses de amostragem, de modo que os pacientes preferem evitar a biópsia [Bedossa 2016, Tapper and Loomba 2018].

Nesse sentido, métodos não invasivos vêm sendo amplamente investigados para avaliação da MASLD [EASL et al. 2024, Santana et al. 2021]. Os mais comuns são a ultrassonografia, ressonância magnética, tomografia computadorizada e a termografia. A tomografia computadorizada utiliza radiação ionizante, enquanto a ressonância magnética apresenta alto custo e menor disponibilidade. A ultrassonografia, apresenta sensibilidade reduzida em casos de esteatose leve e em estágios muito avançados da doença hepática [Cotrim et al. 2016, Powell et al. 2021]. Nesse cenário, a termografia destaca-se por ser um método de baixo custo, não utilizar radiação ionizante e não possuir contraindicações [Farooq and Corcoran 2020, Olaru 2017]. Além disso, a termografia tem sido usada como método preliminar de alterações fisiológicas em diferentes processos patológicos [Lahiri et al. 2012]. Na detecção de doenças hepáticas, estudos indicam que ocorrem alterações na temperatura da superfície do fígado [Farooq and Corcoran 2020].

No âmbito da visão computacional, as Redes Neurais Convolucionais (CNNs, do inglês *Convolutional Neural Networks*) consolidaram-se como uma das principais abordagens em tarefas de classificação de imagens, com aplicações que variam do reconhecimento facial ao diagnóstico médico, onde apresentam desempenho comparável ao de especialistas [Li et al. 2021, Yadav and Jadhav 2019]. Contudo, o campo receptivo local das CNNs limita a captura de dependências espaciais de longo alcance [Khan et al. 2022]. Para superar essa restrição, surgiram arquiteturas baseadas em Transformers, que utilizam mecanismo de autoatenção para modelar relações globais entre diferentes regiões. [Vaswani et al. 2017, Khan et al. 2022]. Um marco dessa evolução foi a Vision Transformer (ViT), que demonstrou resultados competitivos ou superiores às CNNs tradicionais, especialmente quando treinada com grandes conjuntos de dados [Dosovitskiy 2020].

Nesse contexto, ainda são poucos os estudos que utilizam aprendizado profundo aliado a imagens térmicas para classificação da doença hepática. [Pinto et al. 2021] utilizaram técnicas de processamento de imagens aplicadas à imagens térmicas e a CNN AlexNet para classificação da doença hepática, obtendo resultados superiores a 90% em todas as métricas avaliadas. O estudo feito por [Silva et al. 2024] propôs a utilização de uma Rede Neural Siamesa para a classificação de imagens térmicas da região hepática. O trabalho foi motivado pela limitação do número de imagens disponíveis, cenário no qual redes neurais convencionais tendem a apresentar dificuldades de generalização. Os resultados reportados apresentaram acurácia de 71%, precisão de 57% e sensibilidade de 96%. No trabalho apresentado por [Farias et al. 2025], foi empregada uma abordagem utilizando séries temporais em conjunto com a rede TCN, uma Rede Convolucional Temporal, para a classificação de pacientes em grupos com doença hepática e não doentes a partir de imagens termográficas. Os resultados obtidos mostraram-se expressivos alcançando 88% em precisão e 100% em sensibilidade.

Esses estudos evidenciam a viabilidade do uso de imagens térmicas aliadas ao aprendizado profundo, permitindo que os profissionais sejam auxiliados durante o diagnóstico da doença. Os achados da literatura concentram-se em abordagens exclusivamente espaciais ou temporais, apresentando uma lacuna em estudos que investiguem a integração das duas abordagens. Este estudo propõe uma análise comparativa entre as abordagens temporal, espacial e espaço-temporal, para investigar como essa integração influencia na classificação de MASLD, utilizando sequências temporais de termografias.

Este trabalho está estruturado da seguinte forma: a Seção 2 apresenta o método proposto. A Seção 3 apresenta os resultados obtidos e sua discussão. Por fim, a Seção 4 apresenta as conclusões deste trabalho.

2. Materiais e Método

Nesta seção é apresentado o método de classificação das imagens térmicas em dois grupos: doentes (com MASLD) e não doentes. O método é composto por 4 etapas. A etapa 1 consiste na obtenção da base de dados. A etapa 2 compõe a extração da região do fígado das imagens. Na etapa 3 as tabelas de temperatura extraídas são modeladas de acordo com a abordagem utilizada: temporal, espacial ou espaço-temporal, e feito o treinamento da rede utilizada na abordagem. Na etapa 4 é feita a validação de resultados por meio das métricas de avaliação usadas. As etapas são ilustradas por meio da Figura 1.

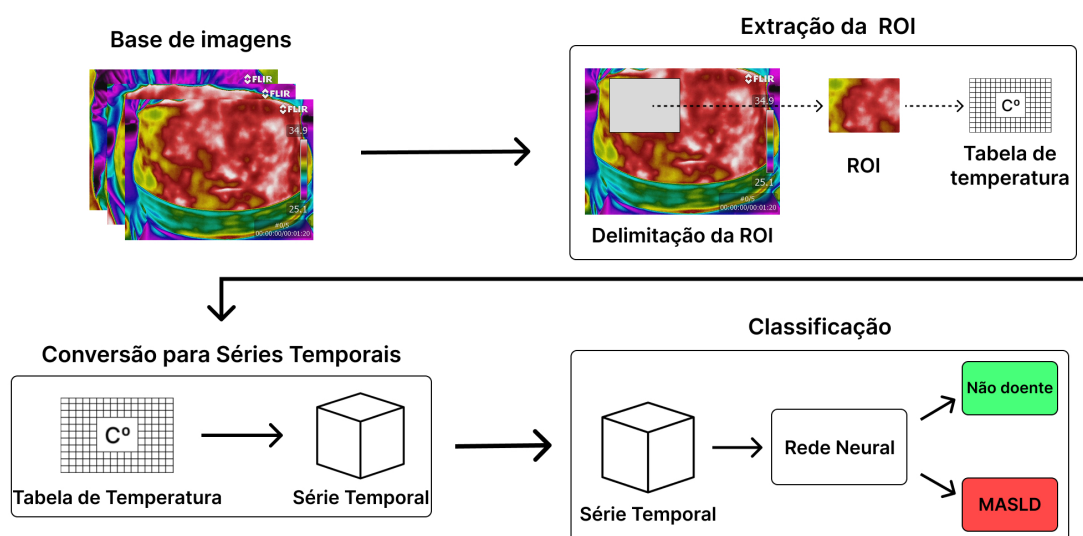


Figura 1. Visão geral do método proposto.

2.1. Base de Dados

A base de dados é composta por 66 pacientes, dos quais 15 foram diagnosticados com MASLD e 51 compõem o grupo não doente. No total, há 75 imagens de pacientes com a doença e 255 imagens de pacientes não doentes. Para aquisição das imagens foi utilizado um termógrafo de sensor infravermelho FLIR - Modelo S650C que possui uma resolução de 640x480 pixels. Os pacientes passam por um período de aclimação no qual permanecem por 15 minutos em um ambiente com 23 °C e umidade relativa de 65%. Eles são posicionados em decúbito dorsal com a área abdominal desnuda durante toda a

aclimatação e captação das imagens. Cinco imagens são capturadas por paciente com um intervalo de vinte segundos entre cada imagem, obtidas com a câmera acima do paciente formando um ângulo de 90°. Esse período visa destacar os padrões irregulares de temperatura na superfície do fígado entre as imagens. O diagnóstico de MASLD foi confirmado por meio de ultrassonografia da região abdominal superior ou total realizada por um especialista. A Figura 2 apresenta exemplos de imagens térmicas utilizadas neste estudo, o retângulo azul na imagem compreende a região afetada pela doença.

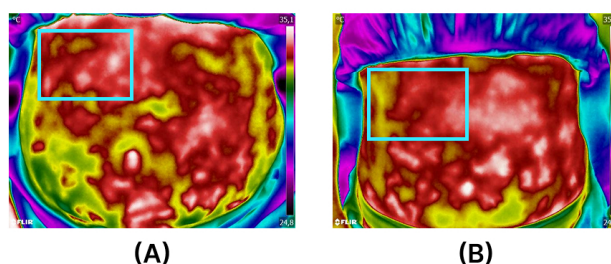


Figura 2. Exemplos de termografias utilizadas: (A) paciente com diagnóstico MASLD e (B) paciente pertencente ao grupo não doente.

2.1.1. Extração da Região de Interesse

Neste trabalho, a região de interesse (ROI, do inglês *Region of Interest*) corresponde à área anatômica que compreende o fígado nas imagens coletadas. [Ozougwu 2017, Brioschi et al. 2006] definem que a região hepática está localizada no quadrante superior direito do abdômen. A delimitação da ROI na região hepática para a classificação das imagens foi feita pois a doença afeta diretamente a região do fígado, gerando padrões térmicos diferentes na mesma. Ao restringir a análise a essa região, reduz-se a influência de áreas irrelevantes da imagem, permitindo que o modelo foque nas características associadas à patologia. A extração das ROIs foi realizada no software FLIR Tools ¹. Com a orientação de um especialista, as dimensões das ROIs foram ajustadas individualmente para cada paciente. Essa variação foi necessária pela variabilidade anatômica e fisiológica dos indivíduos, garantindo que a região hepática fosse delimitada com precisão, independentemente da fisiologia do paciente. As regiões hepáticas foram exportadas como tabelas de temperatura, onde cada posição da matriz contém o valor referente à temperatura, em graus Celsius, de um pixel da imagem original. Para garantir a compatibilidade com a entrada das redes neurais utilizadas, aplicou-se uma técnica de ajuste por corte ou preenchimento (*crop or pad*) em relação ao centro da ROI. Tabelas com dimensões superiores a 224×224 foram centralizadas e recortadas, enquanto tabelas menores receberam preenchimento com valor zero (*zero-padding*) em suas extremidades. Resultando em uma entrada padronizada de 224×224 para todos os modelos.

¹Disponível em: https://flir.custhelp.com/app/answers/detail/a_id/5925/~flir-thermal-studio-suite

2.1.2. Abordagem Temporal

A base de dados utilizada possui simultaneamente estrutura espacial e variação temporal, onde cada imagem possui um intervalo de tempo entre si, buscando capturar alterações térmicas ao decorrer do tempo. Nesse sentido, torna-se necessário definir como essas dimensões serão representadas. Nesta seção, detalha-se a abordagem temporal, que busca representar a evolução dos padrões térmicos.

Para isso, cada tabela foi submetida a um processo de achatamento (*flatten*), resultando em um vetor unidimensional de tamanho 50176 (224×224). A organização sequencial desses vetores para um mesmo paciente gerou uma série temporal de dimensão $N \times 50176$, sendo N o número de passos no tempo e 50176 o resultado do achatamento das tabelas. Embora essa transformação perca consideravelmente as dependências espaciais, ela prioriza a evolução temporal entre as tabelas.

Para utilizar essa sequência temporal, foi utilizada uma Rede Convolutiva Temporal (TCN, do inglês *Temporal Convolutional Networks*). A partir dos princípios das convoluções, [Lea et al. 2017] introduziram um modelo alternativo capaz de explorar dependências temporais em dados visuais, a TCN. A arquitetura utilizada neste trabalho é a *Dilated-TCN*. A Figura 3 apresenta graficamente a montagem das séries temporais após o achatamento das tabelas.

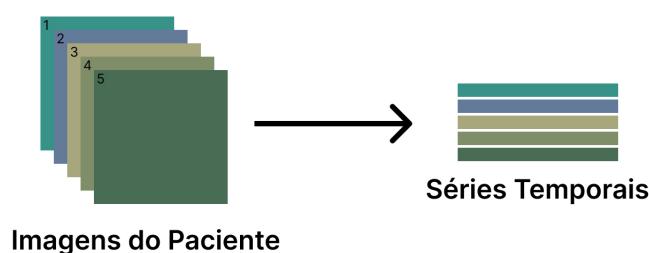


Figura 3. Séries temporais vetorizadas.

2.1.3. Abordagem Espacial com Representação Temporal Multicanal

O processo de achatamento feito na abordagem temporal acaba mitigando a preservação das relações espaciais bidimensionais das tabelas originais. Considerando que a distribuição térmica na região hepática pode conter padrões estruturais relevantes, investigou-se, como estratégia alternativa, uma abordagem que preserve a informação espacial. Embora as arquiteturas utilizadas nesta etapa sejam originalmente projetadas para processamento puramente espacial, adotou-se uma estratégia de empilhamento temporal nos canais de entrada, na qual múltiplos instantes da sequência térmica foram concatenados ao longo da dimensão de canais.

Para isso, os T instantes temporais de cada paciente foram organizados em uma representação multicanal, gerando um tensor de entrada de dimensões $224 \times 224 \times T$. Nessa configuração, a variação temporal é incorporada como diferentes camadas de informações sobrepostas, permitindo que a rede 2D processe simultaneamente

informações de múltiplos instantes sem alterar a estrutura bidimensional original. Assim, diferentemente de modelos espaciais puros da literatura que utilizam quadros isolados, isto é, imagens térmicas individuais correspondentes a um único instante temporal da sequência. A representação multicanal aqui adotada integra o tempo sem descaracterizar a estrutura espacial original, ainda que não seja uma representação independente para a dinâmica temporal.

Para explorar essa dinâmica, foram utilizadas arquiteturas que se destacam tanto no âmbito das CNNs quanto no âmbito das Transformers. As arquiteturas foram a ConvNeXt, com uma proposta que moderniza as CNNs puras com estratégias de macro arquitetura de Transformers [Liu et al. 2022a], e a Multi-Axis Vision Transformer (Max-ViT). A Max-ViT introduz o bloco Max-SA, que combina convoluções com dois tipos diferentes de atenção: atenção em grade e atenção em bloco. Essa abordagem permite que a rede capture dependências locais e globais simultaneamente [Tu et al. 2022]. A montagem desses tensores multicanais é ilustrada na Figura 4.

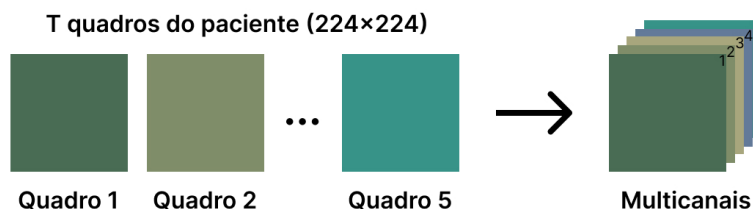


Figura 4. Representação multicanal.

2.1.4. Abordagem Espaço-Temporal

Diferentes das estratégias anteriores, a abordagem espaço-temporal preserva as dimensões espacial e temporal como eixos independentes. Essa estrutura permite que padrões espaciais e suas variações temporais sejam processados de forma conjunta. Nessa formulação, os T quadros são organizados como um volume de formato $T \times C \times 224 \times 224$, mantendo o tempo como uma dimensão própria da entrada, onde T representa o tempo e C é o canal único da matriz térmica.

Para essa abordagem, utilizou-se a Video Swin Transformer, uma arquitetura projetada para análise de dados espaço-temporais. Devido à complexidade computacional quadrática da ViT em relação à resolução das imagens e a ausência de vieses indutivos como localidade, surgiu a Swin Transformer. Essa rede introduziu uma hierarquia através de janelas deslizantes (*shifted windows*), permitindo uma complexidade linear e a possibilidade de modelar tanto dependências locais quanto globais [Liu et al. 2021].

Para a análise de dados espaço-temporais, a Video Swin Transformer adaptou esse conceito para o domínio do vídeo, substituindo as janelas 2D por volumes 3D. O mecanismo de 3D *shifted windows* permite que a rede capture correlações entre quadros adjacentes [Liu et al. 2022b]. Nessa formulação, a arquitetura processa a entrada como um volume espaço-temporal genuíno, diferente do método multicanal usado na abordagem espacial. Esta configuração permite uma extração abrangente de características espaço-

temporais, representando a estratégia mais completa entre as abordagens apresentadas. A Figura 5 apresenta graficamente o modelo espaço-temporal 3D.

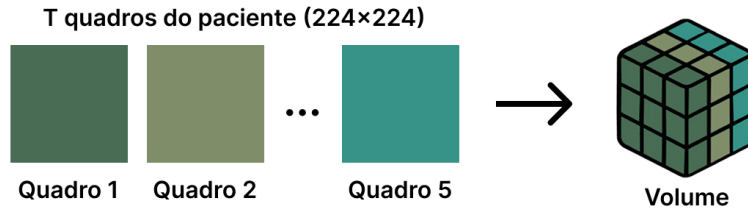


Figura 5. Representação volumétrica 3D.

2.1.5. Classificação por Séries Temporais

Para a classificação, pacientes com MASLD foram considerados como a classe positiva enquanto os não doentes como a classe negativa. O desempenho dos modelos foi avaliado com base em quatro métricas, sendo essas a acurácia, precisão, sensibilidade e F1-Score. Cada uma das métricas é apresentada pelas equações a seguir, onde *VP* significa os verdadeiros positivos, pacientes classificados corretamente com a doença, *FP* os falsos positivos, pacientes incorretamente classificados como doentes, *VN* os verdadeiros negativos, pacientes corretamente classificados como não doentes e *FN* os falsos negativos, pacientes incorretamente classificados como não doentes.

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

$$Precisão = \frac{VP}{VP + FP} \quad (2)$$

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

$$F1 - score = \frac{2 \cdot VP}{2 \cdot VP + FP + FN} \quad (4)$$

A acurácia (Equação 1) é a proporção de acerto geral, sem levar em conta as classes. A precisão (Equação 2) verifica, dentre as classificações de pacientes com a doença realizadas pelo modelo, quantas estão corretas. A sensibilidade (Equação 3) é a proporção de verdadeiros positivos, ou seja, a capacidade de identificar corretamente a doença em pacientes que realmente a possuem. O F1-Score (Equação 4) corresponde a média harmônica entre a precisão e a sensibilidade. Essa métrica é especialmente útil em casos com desbalanceamento entre classes, pois penaliza modelos que apresentam desempenho elevado em apenas uma das métricas.

3. Resultados e Discussão

Esta seção descreve as configurações experimentais adotadas e apresenta o desempenho obtido a partir da aplicação dos modelos propostos. Os resultados são analisados com base nas métricas definidas, comparados entre as diferentes abordagens investigadas e discutidos em relação aos principais achados da literatura.

3.1. Configuração experimental

A base de dados detalhada na Seção 2.1 foi dividida em conjunto de treino (60%), validação (20%) e teste (20%). A divisão foi realizada ao nível de paciente, impedindo que amostras de um mesmo indivíduo fossem distribuídas simultaneamente entre os conjuntos de treino, validação e teste. Esse processo resultou em 40 pacientes no treino, 13 na validação e 13 no teste, correspondendo a 200, 65 e 65 imagens, respectivamente. Como a base utilizada neste trabalho ainda está em construção, a mesma possui poucos dados, resultando em um desbalanceamento entre as classes, com menor número de pacientes pertencentes à classe positiva.

Nesse sentido, adotou-se a técnica de validação cruzada *k-fold cross validation* [Kohavi 1995] conferindo maior robustez estatística e confiabilidade à avaliação do modelo. Esse procedimento garante que todos os pacientes da base foram testados em algum momento, eliminando o viés de uma divisão única e facilitada dos dados. Considerando o desbalanceamento entre as classes, foi utilizada a função de perda *Focal Loss* [Lin et al. 2017]. Os hiperparâmetros utilizados em cada arquitetura foram definidos por meio de uma otimização. O procedimento buscou maximizar o F1-Score, com 20 iterações por modelo. O resultado dos hiperparâmetros estão descritos na Tabela 1.

Tabela 1. Resultado da otimização de hiperparâmetros.

Modelo	LR	Batch Size	Weight Decay	Drop Path	Alpha	Gamma
TCN	1.55×10^{-4}	4	–	–	0.5554	1.50
ConvNeXt	1.84×10^{-5}	4	7.63×10^{-5}	0.2477	0.3619	1.8185
Max-ViT	5.38×10^{-5}	4	2.25×10^{-3}	0.3943	0.3149	4.0791
Video Swin	2.48×10^{-5}	4	3.65×10^{-3}	0.2477	0.3214	2.0658

Todos os modelos foram treinadas com o otimizador *AdamW*. As arquiteturas ConvNeXt, Max-ViT e Video Swin, foram utilizadas suas versões pré-treinadas, otimizando-se apenas parâmetros do treinamento. No caso da TCN, também foram otimizados hiperparâmetros da rede, resultando em uma configuração com 64 filtros, tamanho de kernel 4, dois blocos temporais e esquema de dilatações 1–2–4. O treinamento foi conduzido por no máximo 100 épocas, escolhendo o melhor modelo pelo F1-Score. Os experimentos utilizaram a GPU T4 x2, disponibilizada gratuitamente pela plataforma Kaggle ².

3.2. Resultados dos experimentos

A Tabela 2 apresenta os resultados médios obtidos nos cinco folds e o seu desvio padrão, evidenciando diferenças consistentes entre as abordagens avaliadas. Observa-se que a Video Swin Transformer apresentou o melhor desempenho geral, sendo superior em três

²Disponível em: <https://www.kaggle.com/>

das quatro métricas analisadas, alcançando 79,2% de acurácia, 66% de precisão e 64,6% de F1-Score, além de manter uma boa sensibilidade, alcançando os 73,3%. O ganho em precisão em relação às outras abordagens, indica maior capacidade de reduzir falsos positivos, resultando em um modelo mais equilibrado entre identificação de pacientes doentes e exclusão de pacientes não doentes.

Tabela 2. Resultados médios e desvio padrão obtidos nos experimentos.

Modelo	Acurácia (%)	Precisão (%)	Sensibilidade (%)	F1-Score (%)
TCN	74,4 ± 10,9	45,0 ± 18,7	53,3 ± 26,6	48,0 ± 21,5
ConvNeXt	68,5 ± 13,2	44,0 ± 0,88	73,3 ± 24,8	51,4 ± 0,82
Max-ViT	70,3 ± 23,2	54,8 ± 28,5	80,0 ± 26,6	59,3 ± 21,3
Video Swin	79,2 ± 17,1	66,0 ± 28,7	73,3 ± 24,94	64,6 ± 22,3

As abordagens espaciais com representação multicanal apresentaram sensibilidade expressiva, 73,3% para ConvNeXt e 80% para a Max-ViT, porém baixa precisão, indicando que, embora eficazes na identificação de pacientes doentes, possuem dificuldades em discriminar indivíduos não doentes. O tratamento do tempo como uma composição de canais preserva a estrutura espacial mas desconsidera a ordem e as dependências temporais entre os quadros.

Por outro lado, a abordagem temporal com TCN apresentou baixo desempenho tanto em sensibilidade quanto em precisão, resultando no menor F1-Score entre todos os modelos. Apesar de capturar relações sequenciais, a necessidade de organizar as tabelas térmicas em vetores unidimensionais implica na perda da estrutura espacial original. Nesse contexto, os resultados sugerem que, para a base e a representação dos dados adotadas neste estudo, a topologia espacial pode ter influenciado na distinção das classes.

Em comparação com a literatura, observa-se que os trabalhos de [Farias et al. 2025] e [Pinto et al. 2021], que reportaram métricas superiores a 88%, adotaram protocolos baseados na conversão das tabelas de temperatura em imagens, seguidos de etapas de pré-processamentos e técnicas de aumento de dados (*data augmentation*). Em contraste, este estudo utilizou diretamente tabelas térmicas puras, sem aumento de dados e com um conjunto de dados ampliado em relação aos estudos citados. Essa ampliação implica maior variabilidade biológica e, conseqüentemente, maior desafio para a generalização dos modelos. Assim, embora os resultados absolutos sejam mais moderados, eles reforçam a viabilidade do uso de tabelas de temperatura puras para classificação de MASLD, demonstrando que a preservação da informação térmica, espacial e temporal, é capaz de produzir desempenho consistente.

Em relação ao estudo de [Silva et al. 2024], que também empregou imagens pré-processadas, foram reportadas métricas de 96% em sensibilidade e precisão de 57%. Este comportamento se aproxima dos 54,8% de precisão obtidos pela Max-ViT neste estudo, sugerindo que modelos que utilizam apenas a dimensão espacial podem ter dificuldade de discriminar corretamente os pacientes não doentes. Tal padrão corrobora a hipótese de que a informação espacial, embora relevante, pode não ser suficiente de forma isolada para capturar todos os padrões térmicos associados à MASLD.

Como ilustrado na Figura 6, a distinção visual entre pacientes doentes e não doentes é sutil, evidenciando a dificuldade da tarefa. A análise das amostras revela a comple-

xidade em distinguir características térmicas evidentes que separem os acertos dos erros de classificação. Esse cenário sugere que as decisões do modelo dependem de padrões difíceis, que exigem a integração das dimensões espacial e temporal, reforçando a necessidade de abordagens capazes de capturar relações espaço-temporais mais complexas.

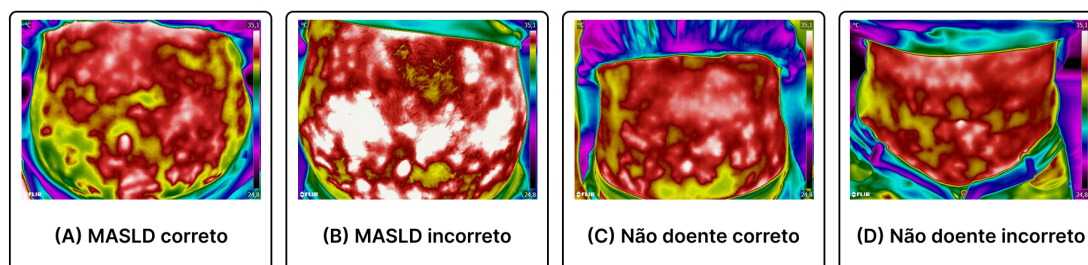


Figura 6. Exemplos de predições realizadas no teste.

As abordagens exclusivamente espaciais ou temporais avaliadas capturam aspectos parciais dos dados, o que se reflete em métricas mais baixas e em menor equilíbrio entre precisão e sensibilidade. Em contrapartida, a abordagem espaço-temporal implementada por meio da Video Swin Transformer apresentou desempenho mais consistente, sugerindo que a assinatura térmica da MASLD envolve dependências espaço-temporais que as análises independentes não conseguem descrever. Dessa forma, os resultados sugerem que a estratégia de integração entre as dimensões espacial e temporal desempenha papel relevante na detecção da doença a partir de tabelas térmicas.

4. Conclusão

Este trabalho investigou diferentes estratégias para a classificação de MASLD a partir de sequências de tabelas térmicas, comparando abordagens espaciais, temporais e espaço-temporais. Os experimentos evidenciaram que a integração das dimensões espacial e temporal influencia na classificação da doença, sendo a abordagem espaço-temporal feita por meio da Video Swin Transformer a que apresentou os resultados mais elevados em três das quatro métricas avaliadas e mais equilíbrio entre precisão e sensibilidade, alcançando 66% e 73,3% respectivamente. Os resultados demonstraram que a utilização dos valores de temperatura da região hepática, aliado a modelos que preservem o tempo e o espaço é um método promissor para a detecção não invasiva da doença hepática.

Como trabalhos futuros, propõe-se a aplicação de estratégias de aumento de dados para fortalecer a generalização dos modelos. Além disso, sugere-se explorar métodos de arquiteturas híbridas que combinem extratores de características espaciais com modelos temporais como CNN-LSTM, ou redes neurais recorrentes como 3D ResNet com GRU, bem como a ampliação da base de dados devido aos poucos dados presentes na atual base.

5. Agradecimentos

O presente trabalho foi realizado com apoio da Coordenação de Aperfeiçoamento de Pessoal de Nível Superior - Brasil (CAPES) - Código de Financiamento 001, Fundação de Amparo a Pesquisa do Maranhão (FAPEMA), Conselho Nacional de Desenvolvimento Científico e Tecnológico (CNPq) e Empresa Brasileira de Serviços Hospitalares (Ebserh) Brazil (Proc. 409593/2021-4).

Referências

- Amini-Salehi, E., Letafatkar, N., Norouzi, N., et al. (2024). Global prevalence of nonalcoholic fatty liver disease: An updated review meta-analysis comprising a population of 78 million from 38 countries. *Archives of Medical Research*, 55(6):103043.
- Bedossa, P. (2016). Histological assessment of naflid. *Digestive Diseases and Sciences*, 61(5):1348–1355.
- Brioschi, M. L., Lin, T. Y., Colman, D., Silva, F. M. R. M., and Teixeira, M. J. (2006). Imaginologia infravermelha no estudo avançado da dor de origem visceral. *Rev Dor: Pesquisa, Clínica e Terapêutica*, 7(4):825–826.
- Cotrim, H. P., Parise, E. R., Figueiredo-Mendes, C., et al. (2016). Nonalcoholic fatty liver disease brazilian society of hepatology consensus. *Arquivos de Gastroenterologia*, 53:118–122.
- Dosovitskiy, A. (2020). An image is worth 16x16 words: Transformers for image recognition at scale. *arXiv preprint arXiv:2010.11929*.
- EASL, EASD, and EASO (2024). Clinical practice guidelines on the management of metabolic dysfunction-associated steatotic liver disease (masld). *Journal of Hepatology*, 81(3):492–542.
- Elshawi, R., Maher, M., and Sakr, S. (2019). Automated machine learning: State-of-the-art and open challenges. *arXiv preprint arXiv:1906.02287*.
- Farias, M. V. S., Silva, I. F. S., Silva, A. C., et al. (2025). Method for detecting non-alcoholic fatty liver disease in abdominal thermography time series using temporal convolutional networks. *Elsevier (To be published)*.
- Farooq, M. A. and Corcoran, P. (2020). Infrared imaging for human thermography and breast tumor classification using thermal images. In *2020 31st Irish Signals and Systems Conference (ISSC)*, pages 1–6. IEEE.
- Friedman, S. L., Neuschwander-Tetri, B. A., Rinella, M., and Sanyal, A. J. (2018). Mechanisms of naflid development and therapeutic strategies. *Nature Medicine*, 24(7):908–922.
- Khan, S. et al. (2022). Transformers in vision: A survey. *ACM Computing Surveys (CSUR)*, 54(10s):1–41.
- Kohavi, R. (1995). A study of cross-validation and bootstrap for accuracy estimation and model selection. In *International Joint Conference on Artificial Intelligence (IJCAI)*, volume 14, pages 1137–1145.
- Lahiri, B. B., Subramaniam, B., Jayakumar, T., and Philip, J. (2012). Medical applications of infrared thermography: A review. *Infrared Physics & Technology*, 55:221–235.
- Lea, C., Flynn, M. D., Vidal, R., Reiter, A., and Hager, G. D. (2017). Temporal convolutional networks for action segmentation and detection. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, pages 156–165.
- Li, Z., Liu, F., Yang, W., Peng, S., and Zhou, J. (2021). A survey of convolutional neural networks: analysis, applications, and prospects. *IEEE Transactions on Neural Networks and Learning Systems*.

- Lin, T. Y., Goyal, P., Girshick, R., He, K., and Dollár, P. (2017). Focal loss for dense object detection. In *Proceedings of the IEEE International Conference on Computer Vision*, pages 2980–2988.
- Liu, Z. et al. (2021). Swin transformer: Hierarchical vision transformer using shifted windows. In *Proceedings of the IEEE/CVF International Conference on Computer Vision*, pages 10012–10022.
- Liu, Z. et al. (2022a). A convnet for the 2020s. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pages 11976–11986.
- Liu, Z. et al. (2022b). Video swin transformer. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pages 3202–3211.
- Olaru, A. (2017). Infrared thermographic evaluation of patients with metastatic vertebral fractures after combined minimal invasive surgical treatment. *The Moldovan Medical Journal*, 60:22–25.
- Ozougwu, J. C. (2017). Physiology of the liver. *International Journal of Research in Pharmacy and Biosciences*, 4(8):13–24.
- Pinto, D. M. et al. (2021). Classificação de esteatose hepática não alcoólica em imagens térmicas da região do fígado utilizando redes neurais convolucionais. In *Simpósio Brasileiro de Computação Aplicada à Saúde (SBCAS)*, pages 302–312. SBC.
- Powell, E. E., Wong, V. W. S., and Rinella, M. (2021). Non-alcoholic fatty liver disease. *The Lancet*, 397(10290):2212–2224.
- Rinella, M. E., Lazarus, J. V., Ratziu, V., Francque, S. M., et al. (2023). A multisociety delphi consensus statement on new fatty liver disease nomenclature. *Hepatology*, 78(6):1966–1986.
- Santana, J. T., Mota, A. V. H., Gonzaga, Y. H. G., et al. (2021). Perfil metabólico e antropométrico dos pacientes obesos e não obesos portadores de esteatose hepática não alcoólica. *Revista Eletrônica Acervo Saúde*, 13(2):e5525.
- Silva, M. P., Silva, A. C., and de Paiva, A. C. (2024). Classification of non-alcoholic fatty liver disease in thermal images of the liver using a siamese neural network. In *Brazilian Conference on Intelligent Systems*, pages 260–269. Springer Nature Switzerland.
- Tapper, E. B. and Loomba, R. (2018). Noninvasive imaging biomarker assessment of liver fibrosis by elastography in nafld. *Nature Reviews Gastroenterology & Hepatology*, 15(5):274–282.
- Tu, Z. et al. (2022). Maxvit: Multi-axis vision transformer. In *European Conference on Computer Vision*, pages 459–479. Springer Nature Switzerland.
- Vaswani, A. et al. (2017). Attention is all you need. *Advances in Neural Information Processing Systems*, 30.
- Yadav, S. S. and Jadhav, S. M. (2019). Deep convolutional neural network based medical image classification for disease diagnosis. *Journal of Big Data*, 6(1):1–18.
- Zuluaga-Gomez, J., Masry, Z. A., Benagoune, K., et al. (2021). A cnn-based methodology for breast cancer diagnosis using thermal images. *Computer Methods in Biomechanics and Biomedical Engineering: Imaging & Visualization*, 9(2):131–145.