When Do Spectral-Spatial Transformers Surpass Classical Ensembles in Hyperspectral Image Classification? A Leakage-Aware Supervised Evaluation

João Vítor Fonseca de Mendonça[1], Anne Magaly de Paula Canuto[1]
Marcio Eduardo Kreutz[1]

[1]Departamento de Informática e Matemática Aplicada (DIMAp)
Universidade Federal do Rio Grande do Norte (UFRN)
Natal – RN – Brazil

joao.vitor.081@ufrn.edu.br, anne.canuto@ufrn.br
kreutz@dimap.ufrn.br

Abstract. *Hyperspectral image classification is strongly affected by limited labels, class imbalance, and spatial autocorrelation. These properties make comparisons between classical classifiers and Transformer-based models particularly sensitive to how train and test samples are separated. We evaluate spectral and spectral-spatial Transformers together with classical supervised baselines on five public scenes, using five seeds and leakage-aware spatial partitions. The main analysis contrasts SVM-RBF with a spectral-spatial Transformer under random and spatially blocked protocols. The Transformer reaches very high overall accuracy under random pixel splits, but its advantage is reduced or reversed when evaluation is spatially separated. The behavior also depends on the metric: gains in overall accuracy do not always translate into better average accuracy or macro-F1. These results show that the practical value of spatial context depends on the evaluation protocol, the scene, and the target metric, while classical models remain strong alternatives when geographic generalization and computational cost are important.*

Resumo. *A classificação de imagens hiperespectrais é fortemente afetada por poucos rótulos, desbalanceamento de classes e autocorrelação espacial. Essas características tornam a comparação entre classificadores clássicos e modelos baseados em Transformers particularmente sensível à forma como treino e teste são separados. Avaliamos Transformers espectrais e espectral-espaciais em conjunto com modelos clássicos em cinco cenas públicas, com cinco sementes e partições espaciais que controlam vazamento. A análise principal contrasta SVM-RBF com um Transformer espectral-espacial sob protocolos aleatório e espacialmente bloqueado. O Transformer alcança acurácia global muito alta na divisão aleatória por pixel, mas sua vantagem diminui ou se inverte quando a avaliação separa regiões espacialmente. O comportamento também depende da métrica: ganhos de acurácia global nem sempre produzem melhores valores de acurácia média ou F1 macro. Os resultados mostram que o valor prático do contexto espacial depende do protocolo de avaliação, da cena e da métrica de interesse, enquanto modelos clássicos permanecem alternativas fortes quando generalização geográfica e custo computacional são relevantes.*

1. Introdução

Imagens hiperespectrais (HSI) representam cada pixel por dezenas ou centenas de bandas espectrais contíguas. Essa riqueza permite discriminar materiais e coberturas do solo que seriam indistinguíveis em imagens RGB, sendo importante em agricultura de precisão, monitoramento ambiental e mapeamento urbano. Em contrapartida, a classificação supervisionada de HSI enfrenta alta dimensionalidade, redundância entre bandas, escassez de rótulos e desbalanceamento entre classes [Camps-Valls et al. 2014, Chen et al. 2016].

Classificadores clássicos, como SVM-RBF e Random Forest (RF), permanecem modelos de referência particularmente fortes nesse regime de poucos dados. RF e outros ensembles de árvores são robustos, paralelizáveis e relativamente baratos [Belgiu and Draguţ 2016]. Por outro lado, redes profundas e Transformers podem aprender representações espectrais e espaciais diretamente. O SpectralFormer explora a estrutura sequencial das bandas [Hong et al. 2022], enquanto o Hyperspectral Image Transformer (HiT) combina contexto local e atenção para modelar dependências espectral-espaciais [Yang et al. 2022].

Entretanto, a afirmação de que Transformers são superiores não deve ser aceita sem qualificações. Modelos com atenção possuem maior capacidade e custo computacional, enquanto HSI tipicamente oferece poucos exemplos rotulados. Além disso, a prática comum de dividir pixels aleatoriamente pode posicionar amostras espacialmente adjacentes em treino e teste. Para modelos baseados em patches, a vizinhança de um pixel de teste pode se sobrepor direta ou indiretamente à informação usada no treinamento, produzindo estimativas excessivamente otimistas de generalização [Hu et al. 2023]. Revisões recentes também identificam escassez de rótulos, custo computacional e generalização como desafios abertos da classificação HSI baseada em Transformers [Cao et al. 2026].

A questão central deste artigo é, portanto, mais específica do que perguntar qual arquitetura atinge a maior acurácia: sob quais condições Transformers espectrais e espectral-espaciais oferecem vantagem prática sobre modelos clássicos em HSI? A análise separa desempenho sob divisão aleatória e sob divisão espacialmente bloqueada, considera métricas globais e macro e explicita o custo de treinamento. As contribuições do trabalho são:

- propor um protocolo supervisionado que separa explicitamente entrada espectral de entrada espectral-espacial e compara divisão aleatória com divisão por blocos espaciais e buffer;
- avaliar famílias clássicas e neurais sob o mesmo desenho experimental, com atenção especial ao contraste entre SVM-RBF e Transformer espectral-espacial;
- apresentar resultados com múltiplas sementes em cinco cenas públicas, evidenciando cenários em que o contexto espacial ajuda e cenários em que modelos clássicos permanecem competitivos;
- registrar sementes, partições, parâmetros, predições e métricas para favorecer reprodutibilidade e auditoria experimental.

2. Trabalhos Relacionados

2.1. Classificação clássica e profunda de HSI

SVM, k-NN e RF formam a base de grande parte da literatura clássica de HSI. Em especial, RF combina amostragem bootstrap e subespaços aleatórios de atributos, reduzindo a variância de árvores individuais [Belgiu and Draguţ 2016]. Tais métodos usam tradicionalmente apenas o vetor espectral central e, por isso, constituem uma referência importante para medir o valor adicional do contexto espacial.

Redes profundas passaram a explorar simultaneamente informação espacial e espectral por meio de CNNs 1D, 2D e 3D. Embora possam melhorar a separação de classes similares, esses modelos também sofrem com poucos rótulos e com a possibilidade de memorizar estruturas locais [Chen et al. 2016]. O problema se torna especialmente relevante quando patches são avaliados sobre divisões aleatórias de pixels.

2.2. Transformers e vazamento espacial

Transformers oferecem mecanismos de atenção adequados para dependências de longo alcance. SpectralFormer trata bandas como uma sequência de tokens e procura capturar relações espectrais não locais [Hong et al. 2022]; HiT incorpora operações convolucionais e atenção para combinar discrepâncias espectrais e contexto espacial [Yang et al. 2022]. Essas arquiteturas, contudo, não dispensam avaliação cuidadosa.

A revisão de [Hu et al. 2023] mostra que autocorrelação espacial e participação indireta de informação de teste podem elevar artificialmente resultados de classificadores profundos de HSI. O presente estudo torna esse problema parte explícita da pergunta experimental. A divisão aleatória é mantida para comparabilidade com parte da literatura, enquanto a divisão por blocos espaciais, com faixa de exclusão proporcional ao raio do patch, é tratada como avaliação mais exigente de generalização geográfica.

3. Estudo Motivador e Hipóteses

Um estudo preliminar em Indian Pines avaliou quatro representações espectrais: todas as 200 bandas; amostragem estratificada de 50%; seleção por correlação de Pearson; e PCA com 95% da variância. Foram considerados k-NN, árvore de decisão, GaussianNB e MLP, seguidos por Bagging, AdaBoost, RF, Stacking e XGBoost. RF foi o melhor ensemble médio (81,49% em validação cruzada sobre as quatro representações), enquanto XGBoost foi próximo (81,18%). Na base original, RF com 100 árvores atingiu 88,15% em validação cruzada; MLP foi o melhor classificador individual espectral relatado, com 85,95% em hold-out 70/30.

Posteriormente, descritores HOG/LBP em patches produziram resultados próximos de 100%. Uma auditoria do código mostrou que PCA, normalização e extração de patches foram realizados antes da divisão treino/teste. Portanto, patches de pixels adjacentes podiam compartilhar informação entre partições. Esses valores não são usados como referência quantitativa neste artigo; eles motivam uma reavaliação com controle explícito de vazamento.

Definimos três hipóteses falsificáveis. H1: a vantagem relativa de Transformers espectral-espaciais aumenta com a disponibilidade de rótulos e a heterogeneidade espacial, mas não é necessariamente positiva em todas as cenas. H2: com poucos rótulos

ou divisão espacial estrita, RF, XGBoost e SVM-RBF são competitivos ou superiores a Transformers maiores. H3: a diferença entre resultados aleatórios e bloqueados tende a ser maior para modelos de patch do que para modelos estritamente espectrais.

4. Conjuntos de Dados e Protocolo Experimental

4.1. Cenas, domínios e critérios de inclusão

O conjunto experimental foi construído para não depender de uma única cena agrícola. Indian Pines, Salinas, Botswana e KSC representam vegetação, cultivos e áreas naturais; Pavia University representa ambiente urbano. As cenas públicas de Indian Pines, Salinas, Pavia, KSC e Botswana são disponibilizadas pelo GIC/UPV-EHU [Grupo de Inteligencia Computacional 2026]. Essa seleção permite contrastar cenas agrícolas, urbanas e naturais sob protocolos de divisão aleatória e espacialmente bloqueada.

A diversidade não é inferida apenas a partir do número de pixels rotulados. Para cada cena, consideramos domínio, resolução espacial, número de bandas úteis, frequência por classe e distribuição de rótulos. Essa caracterização ajuda a separar o efeito de maior volume de dados de mudanças de domínio ou desbalanceamento.

Tabela 1. Conjunto de dados avaliado. R: divisão aleatória; B: divisão espacial bloqueada.

Cena	Domínio	Linhas × colunas	Bandas	Classes	Protocolo
Indian Pines	Agrícola	145×145	200	16	R, B
Salinas	Agrícola	512×217	204	16	B
Pavia University	Urbano	610×340	103	9	B
KSC	Áreas úmidas	512×614	176	13	R
Botswana	Natural/agrícola	1476×256	145	14	R, B

4.2. Partições sem vazamento e orçamento de rótulos

As cenas são avaliadas sob um ou ambos os protocolos indicados na Tabela 1. Em R, pixels rotulados são divididos de forma estratificada em 60/20/20 para treino, validação e teste. Como a seleção é feita no nível do pixel, amostras de treino e teste podem ficar espacialmente próximas; portanto, R mede principalmente interpolação dentro da mesma cena e mantém comparabilidade com parte da literatura. Em B, a imagem é subdividida em blocos espaciais; os blocos são atribuídos a treino, validação ou teste; e centros cujo patch cruza uma fronteira são removidos. Assim, B aproxima uma situação de generalização regional: o modelo é treinado em algumas áreas da imagem e avaliado em regiões espacialmente separadas.

Para patch p × p, o buffer mínimo é $\lfloor p/2 \rfloor$, de modo que nenhuma janela centrada em uma amostra válida atravesse duas partições. O gerador de B cria múltiplos candidatos por semente e seleciona a partição com maior cobertura de classes após a aplicação do

buffer. Essa verificação é indispensável: em Indian Pines, classes extremamente raras não ocorrem em todas as regiões válidas, o que torna AA e F1 macro uma análise de sensibilidade nesse conjunto. Salinas e Pavia University preservaram todas as classes na rodada com múltiplas sementes bloqueada.

Os experimentos usam aproximadamente 60% dos pixels rotulados para treinamento, 20% para validação e 20% para teste antes da aplicação do filtro espacial. Todos os índices são definidos antes do pré-processamento. Normalização, PCA e qualquer seleção de bandas são ajustadas exclusivamente no treino de cada repetição. Essa ordem impede que estatísticas ou componentes do teste informem o modelo durante o treinamento.

4.3. Modelos, treinamento e reprodutibilidade

Os modelos espectrais de referência são RF, XGBoost, SVM-RBF e MLP. RF e XGBoost representam ensembles de árvores; SVM-RBF é incluído por ser um baseline clássico forte em HSI; MLP isola o efeito de uma rede densa sem atenção. Esses modelos recebem apenas o vetor espectral central. XGBoost foi estabilizado localmente no método approx, com 20 árvores, devido a limitações observadas no ambiente macOS usado nos experimentos.

Dois Transformers compactos são avaliados. O Transformer espectral tokeniza grupos contíguos de bandas e aplica autoatenção ao longo do eixo espectral. O Transformer espectral-espacial recebe patches 9×9 em uma representação PCA de 20 componentes ajustada exclusivamente no treino e aplica atenção sobre tokens espaciais.

As configurações são mantidas fixas por família de modelo dentro de cada conjunto de experimentos e avaliadas nas mesmas partições. Para redes, usamos inicializações controladas, early stopping quando aplicável e pesos de classe quando justificado pelo desbalanceamento. Cada execução registra semente, manifesto de partição, versões de bibliotecas, número de parâmetros, tempo de treino e tempo de inferência. O objetivo é avaliar o compromisso entre desempenho e custo, e não apenas o maior valor de OA.

4.4. Métricas e agregação multi-seed

As métricas primárias são acurácia global (OA), acurácia média por classe (AA) e F1 macro. Kappa, F1 por classe, matrizes de confusão e mapas de erro são métricas complementares previstas pelo protocolo. OA isolada é insuficiente em cenas com forte desbalanceamento, pois uma classe majoritária pode dominar a pontuação sem representar a qualidade média do mapeamento.

A unidade de comparação é a tupla conjunto de dados, protocolo, modelo e semente. Reportamos média e desvio-padrão em cinco sementes e diferenças em pontos percentuais entre pares de modelos de interesse. Essa agregação multi-seed reduz a dependência de uma única partição favorável e permite avaliar simultaneamente tendência central e variabilidade entre execuções. Não são empregados testes de hipótese ou critérios de significância: as comparações são descritivas e interpretadas em conjunto com o protocolo, a cobertura de classes e o custo computacional.

4.5. Critérios de equidade

Uma comparação entre famílias de modelos só é informativa quando a modalidade de entrada é explicitada. RF, XGBoost, SVM-RBF, MLP e Transformer espectral usam ape-

nas o vetor espectral central. O Transformer espectral-espacial é analisado como uma categoria distinta, pois recebe informação adicional de vizinhança. Essa separação evita atribuir exclusivamente à arquitetura um ganho que pode decorrer, em parte, do acesso ao contexto espacial.

Tabela 2. Configurações comparadas e modalidade de entrada.

Modelo	Entrada	Controle principal
Random Forest	Espectro central	Árvores, atributos por divisão
XGBoost	Espectro central	Árvores, profundidade, taxa de aprendizado
SVM-RBF	Espectro central	C e γ
MLP	Espectro central	Largura, regularização, early stopping
Transformer espectral	Tokens de bandas	Largura, camadas, grupos espectrais
Transformer espectral-espacial	Patch + PCA treinada	Patch, tokens, largura, camadas

Outra fonte de assimetria é a seleção pós-teste. Para evitá-la, cada família usa uma validação interna fixa em cada semente, e o teste é consultado apenas após congelar a configuração avaliada. Essa decisão é especialmente importante para Transformers, pois ajustar repetidamente profundidade, largura ou número de épocas olhando a métrica de teste pode criar uma vantagem artificial sobre os modelos clássicos.

4.6. Reprodutibilidade e rastreabilidade

O experimento é organizado para que as tabelas possam ser reconstruídas a partir dos arquivos salvos. Cada linha de resultado contém conjunto de dados, protocolo, semente, modelo, métricas, tempo e metadados da partição. Os índices de treino, validação e teste são serializados; os dados brutos permanecem separados do código; e os scripts registram versões de Python, bibliotecas e dispositivo de execução. O código também registra explicitamente a limitação do backend XGBoost em macOS, evitando que uma alteração silenciosa de método de árvore seja interpretada como diferença algorítmica.

5. Resultados com Múltiplas Sementes

Esta seção apresenta a avaliação com cinco sementes (42–46) e orçamento de treinamento de aproximadamente 60% dos pixels rotulados. Indian Pines, Salinas e Pavia University usaram três épocas máximas nos experimentos compactos com Transformers, enquanto Botswana e KSC usaram 20 épocas. Como os orçamentos de otimização diferem entre grupos de experimentos, comparações absolutas entre conjuntos de dados são interpretadas com cautela; as comparações centrais são feitas entre modelos avaliados sob a mesma partição e o mesmo orçamento dentro de cada cena.

A análise quantitativa principal concentra-se em SVM-RBF e no Transformer espectral-espacial (T-SS). Entre as configurações avaliadas, esse par produz o contraste

mais informativo para a pergunta central do trabalho, pois opõe um classificador espectral clássico forte a um modelo que incorpora explicitamente a vizinhança espacial. RF, XGBoost, MLP e o Transformer estritamente espectral foram mantidos no protocolo como referências de controle, mas não alteraram o padrão central discutido nesta seção; por isso, seus resultados não são reproduzidos na tabela principal. Essa escolha restringe a comparação apresentada aos dois modelos que melhor evidenciam o efeito do protocolo e do contexto espacial, sem implicar um ranking global entre todas as famílias.

Tabela 3. Resultados com cinco sementes: média ± desvio-padrão. T-SS: Transformer espectral-espacial. R: divisão aleatória; B: divisão espacial bloqueada. Valores em %.

Cena	Partição/Modelo	OA	AA	F1 macro
Indian Pines[†]	R / SVM-RBF	88,38 ± 0,75	90,90 ± 0,94	89,47 ± 1,11
Indian Pines[†]	R / T-SS	92,46 ± 1,54	75,56 ± 1,63	75,68 ± 1,40
Indian Pines[†]	B / SVM-RBF	55,40 ± 8,05	41,04 ± 3,63	38,68 ± 4,02
Indian Pines[†]	B / T-SS	59,23 ± 8,75	49,06 ± 8,47	48,77 ± 7,67
Salinas	B / SVM-RBF	87,94 ± 3,24	94,80 ± 1,11	94,84 ± 1,13
Salinas	B / T-SS	90,52 ± 4,04	93,78 ± 3,51	92,84 ± 4,30
Pavia University	B / SVM-RBF	69,10 ± 3,01	85,55 ± 1,42	70,72 ± 2,33
Pavia University	B / T-SS	58,83 ± 5,41	87,90 ± 3,09	73,62 ± 4,65
Botswana	R / SVM-RBF	95,06 ± 0,33	95,73 ± 0,29	95,53 ± 0,36
Botswana	R / T-SS[‡]	99,85 ± 0,19	99,87 ± 0,16	99,85 ± 0,18
Botswana	B / SVM-RBF[§]	75,52 ± 0,00	79,42 ± 0,00	78,00 ± 0,00
Botswana	B / T-SS[‡§]	57,86 ± 8,80	66,39 ± 5,06	56,77 ± 5,67
KSC	R / SVM-RBF	94,93 ± 0,56	92,37 ± 0,64	92,00 ± 0,81
KSC	R / T-SS[‡]	99,58 ± 0,32	99,22 ± 0,55	99,24 ± 0,56

[†] Em Indian Pines, classes extremamente raras não ocorrem em todas as regiões bloqueadas após o buffer; os valores macro são análise de sensibilidade. [‡] Botswana e KSC usam 20 épocas; as demais rodadas T-SS usam três. [§] Em Botswana-B, blocos de 21 pixels preservam as 14 classes após o buffer.

Sob R, T-SS obteve maior OA nas cenas avaliadas, mas em Indian Pines sua AA e F1 macro ficaram abaixo de SVM-RBF. Em Botswana, após 20 épocas, T-SS atingiu 99,85% de OA e F1 macro, superando SVM-RBF por 4,79 e 4,32 pontos percentuais, respectivamente. KSC reproduz esse padrão em uma cena de áreas úmidas: T-SS supera SVM-RBF por 4,65 pontos de OA e 7,24 pontos de F1 macro. Esses resultados, contudo, devem ser lidos dentro do orçamento de treinamento específico dessas cenas.

Sob B, a interpretação muda. Em Botswana, SVM-RBF obtém 75,52% de OA e 78,00% de F1 macro, enquanto T-SS alcança 57,86% e 56,77%. Em Salinas, SVM-RBF apresenta maior AA e F1 macro, enquanto T-SS obtém maior OA (90,52% contra 87,94%) a custo computacional aproximadamente 24 vezes maior. Em Pavia University, SVM-RBF obtém maior OA, mas T-SS apresenta maior AA e F1 macro. Assim, a escolha do modelo depende da métrica, do desbalanceamento, do orçamento de treinamento e, sobretudo, do protocolo de divisão.

5.1. Efeito do protocolo de divisão

O resultado mais marcante é a queda observada quando a avaliação passa de R para B. Em Indian Pines, a OA média de T-SS muda de 92,46% no protocolo aleatório para 59,23% no bloqueado. Essa diferença não pode ser atribuída exclusivamente a vazamento, pois B também altera o tamanho efetivo e a distribuição espacial do conjunto de treino. Ainda assim, o contraste mostra que a estratégia de partição é um fator experimental indispensável e justifica tratar a avaliação bloqueada como referência mais exigente de generalização geográfica.

Indian Pines também ilustra uma armadilha da OA isolada. O modelo espacial obtém maior OA sob R, mas SVM-RBF apresenta AA e F1 macro superiores. A diferença entre OA e métricas macro mostra que o ganho global não é distribuído de forma uniforme entre as classes. Em outras palavras, uma melhora em OA pode coexistir com pior desempenho médio por classe, o que reforça a necessidade de interpretar as três métricas em conjunto.

5.2. Interpretação por tipo de cena

Salinas e Pavia University oferecem regimes complementares. Em Salinas, cena agrícola com grandes talhões, SVM-RBF preserva AA e F1 macro superiores sob B. Nesse cenário, a assinatura espectral já fornece separação forte entre várias classes, e o ganho de OA de T-SS não se traduz em melhora das métricas macro. Assim, o contexto espacial acrescenta informação, mas não substitui a leitura por classe quando há desbalanceamento.

Pavia University representa um ambiente urbano com objetos menores e maior alternância espacial entre materiais. Sob B, T-SS eleva AA e F1 macro mesmo sem vencer em OA. O resultado sugere que o contexto local é útil para parte das classes mesmo quando não aumenta a OA. Como o efeito aparece nas métricas macro, ele é mais compatível com uma melhora distribuída entre classes do que com um ganho concentrado apenas nas classes majoritárias.

5.3. Custo e decisão prática

O tempo de treino mediano revela um compromisso relevante. Em Salinas bloqueado, SVM-RBF treinou em aproximadamente 0,52 s, enquanto T-SS consumiu cerca de 13,14 s. Em Pavia University, o contraste foi de 0,28 s contra 8,90 s. Esses valores dependem do hardware e excluem ajuste extensivo de hiperparâmetros, mas mostram que a maior acurácia não deve ser o único critério de escolha.

Para aplicações com orçamento computacional restrito e prioridade para classificação espectral rápida, SVM-RBF constitui uma alternativa forte nos cenários avaliados. Quando métricas macro e contexto espacial têm maior peso, T-SS pode ser vantajoso em determinadas cenas, desde que o custo adicional seja aceitável. Por isso, a escolha entre os modelos depende do objetivo da aplicação: velocidade e simplicidade favorecem SVM-RBF, enquanto T-SS se torna mais atraente quando o contexto espacial compensa o custo adicional.

6. Análise de Cenários

Os resultados mostram que o efeito do contexto espacial muda com a cena e com a forma de separar treino e teste. Em Salinas, o SVM-RBF mantém vantagem em AA e F1 macro

sob divisão bloqueada, apesar da maior OA de T-SS. Em Pavia University ocorre o inverso nas métricas macro: o T-SS melhora AA e F1 mesmo com OA menor. O contraste indica que o benefício do contexto espacial não aparece da mesma forma em todos os domínios nem em todas as métricas.

KSC e Botswana ampliam a análise para além do contraste agricultura-cidade. Nas duas cenas, T-SS apresenta resultados muito altos sob divisão aleatória quando recebe 20 épocas. Em Botswana, entretanto, a divisão bloqueada com cobertura explícita das 14 classes inverte o ranking e favorece SVM-RBF. A inversão é importante porque mostra que a vantagem obtida com pixels aleatoriamente misturados entre treino e teste não se transfere necessariamente para regiões espacialmente separadas.

Também é necessário separar volume de dados de diversidade efetiva. Muitos pixels provenientes de uma mesma classe ou região podem aumentar o tamanho nominal do conjunto sem produzir exemplos independentes. Por isso, as conclusões são formuladas em função de domínio, número de classes, equilíbrio, resolução espacial, sensor e protocolo. O estudo não pretende extrapolar os resultados para todos os cenários de HSI, mas identificar padrões que merecem ser testados em avaliações cross-scene e em conjuntos adicionais.

7. Ameaças à Validade

Há cinco limitações que devem ser consideradas. Primeiro, o orçamento de treinamento dos Transformers não é uniforme: Botswana e KSC usam 20 épocas, enquanto Indian Pines, Salinas e Pavia University usam três. Assim, comparações absolutas entre conjuntos de dados também refletem diferenças de otimização. Segundo, em Indian Pines, classes muito raras não aparecem em todas as regiões válidas após a aplicação do buffer; por isso, AA e F1 macro no protocolo bloqueado devem ser lidas como análise de sensibilidade. Terceiro, XGBoost foi executado no modo `approx` com 20 árvores por limitações do ambiente macOS, o que reduz sua força como representante de boosting. Quarto, a Tabela 3 destaca SVM-RBF e T-SS, escolhidos por formarem o contraste mais informativo para a questão central; os demais modelos do protocolo não são mostrados em detalhe e, portanto, não sustentamos um ranking global entre as seis famílias. Quinto, matrizes de confusão e mapas de erro não são apresentados nesta versão, o que limita a análise espacial e por classe dos erros.

Esses pontos não alteram o contraste principal observado entre divisão aleatória e divisão bloqueada, mas limitam sua generalização. Em particular, os resultados devem ser entendidos como evidência sobre as cenas e configurações avaliadas, e não como prova de superioridade universal de uma família de modelos.

8. Conclusão

Este trabalho avaliou em quais condições Transformers espectral-espaciais oferecem vantagem prática sobre modelos clássicos na classificação supervisionada de imagens hiperespectrais. O resultado mais consistente não é a superioridade de uma arquitetura, mas a forte dependência do protocolo de avaliação. Sob divisão aleatória por pixel, T-SS alcança OA muito alta e supera SVM-RBF em várias cenas. Quando treino e teste são separados espacialmente, essa vantagem diminui e, em Botswana, chega a se inverter.

As métricas também mudam a conclusão. Em Indian Pines, T-SS apresenta maior OA no protocolo aleatório, enquanto SVM-RBF mantém AA e F1 macro superiores. Em Salinas bloqueado, SVM-RBF conserva vantagem nas métricas macro; em Pavia University, T-SS melhora AA e F1 macro mesmo com OA inferior. Portanto, escolher o modelo apenas pela acurácia global pode ocultar diferenças importantes entre classes.

Do ponto de vista prático, SVM-RBF permanece uma alternativa forte quando custo, simplicidade e generalização espacial têm maior peso. T-SS é mais atraente quando o contexto local acrescenta informação útil e o custo de treinamento é aceitável. O principal cuidado metodológico é evitar que resultados obtidos com divisão aleatória sejam interpretados diretamente como desempenho em regiões geograficamente novas.

Como continuidade, o protocolo pode ser ampliado com orçamento de otimização uniforme, apresentação completa dos modelos de controle, matrizes de confusão, mapas de erro e avaliações entre cenas e sensores. Essas extensões permitiriam verificar se os padrões observados neste estudo se mantêm quando a generalização exige transferência espacial mais forte.

Referências

[Belgiu and Draguţ, 2016] Belgiu, M. and Draguţ, L. (2016). Random forest in remote sensing: A review of applications and future directions. *ISPRS Journal of Photogrammetry and Remote Sensing*, 114:24–31.

[Camps-Valls et al., 2014] Camps-Valls, G., Tuia, D., Bruzzone, L., and Benediktsson, J. A. (2014). Advances in hyperspectral image classification: Earth monitoring with statistical learning methods. *IEEE Signal Processing Magazine*, 31(1):45–54.

[Cao et al., 2026] Cao, Y. et al. (2026). Transformers meet hyperspectral imaging: A comprehensive study of models, challenges and open problems. *Engineering Applications of Artificial Intelligence*, 169:113947.

[Chen et al., 2016] Chen, Y., Jiang, H., Li, C., Jia, X., and Ghamisi, P. (2016). Deep feature extraction and classification of hyperspectral images based on convolutional neural networks. *IEEE Transactions on Geoscience and Remote Sensing*, 54(10):6232–6251.

[GIC, 2026] Grupo de Inteligencia Computacional (2026). Hyperspectral remote sensing scenes. Universidad del País Vasco (UPV/EHU). Accessed: 2026-06-22.

[Hong et al., 2022] Hong, D., Han, Z., Yao, J., Gao, L., Zhang, B., and Chanussot, J. (2022). SpectralFormer: Rethinking hyperspectral image classification with transformers. *IEEE Transactions on Geoscience and Remote Sensing*, 60:1–15.

[Hu et al., 2023] Hu, Y. et al. (2023). Information leakage in deep learning-based hyperspectral image classification: A survey. *Remote Sensing*, 15(15):3793.

[Yang et al., 2022] Yang, X., Cao, W., Yao, L., and Zhou, Y. (2022). Hyperspectral image transformer classification networks. *IEEE Transactions on Geoscience and Remote Sensing*, 60:1–15.