

Análise e classificação do sistema agroflorestal cabruca com aprendizagem de máquina no Bioma Mata Atlântica – Brasil

Analysis and classification of the cabruca agroforestry system with machine learning in the Atlantic Forest Biome – Brazil

Vinícius de Amorim Silva *, **Hercules da Silva Carvalho ****, **Paulo Sérgio Vila Nova Souza *****

* Centro de Formação em Tecnociências e Inovação, Universidade Federal do Sul da Bahia - UFSB, vinicius@ufsb.edu.br

** Centro de Formação em Tecnociências e Inovação, Universidade Federal do Sul da Bahia – UFSB, hercules.carvalho@cja.ufsb.edu.br

*** Centro de Formação em Ciências Agroflorestais, Universidade Federal do Sul da Bahia - UFSB, paulosvn@gfe.ufsb.edu.br

<http://dx.doi.org/10.5380/raega.v64i1.96730>

Resumo

A ausência de rotulagem e similaridade espectrais dos sistemas cabruca no bioma Mata Atlântica para as bacias hidrográficas dos rios Almada, Cachoeira e Una nos bancos de dados dificulta a classificação adequada dessas áreas. Assim, este trabalho objetiva analisar a classificação das áreas de cabruca nas referidas bacias por meio do uso de imagens de diferentes resoluções e aprendizado de máquina. A metodologia inclui: a determinação das classes de cobertura e uso da terra, a obtenção de imagens de satélite, o processamento das imagens com algoritmos em *Python* e a avaliação da classificação supervisionada. Os resultados mostram que os algoritmos de recorte são eficazes no processamento das imagens de satélite, preservando informações geográficas essenciais. Além disso, o modelo de Redes Neurais alcançou uma acurácia de 92%, independente da diversidade de resoluções espaciais. No entanto, recomenda-se usar os resultados da classificação das áreas de cabruca como estimativas, devido à dificuldade dos algoritmos em distinguir essas áreas da Floresta Ombrófila Densa.

Palavras-chave:

Inteligência Artificial, Geoprocessamento, Cacau sombreado, Conservação da biodiversidade.

Abstract

The absence of labeling and spectral similarity of cabruca systems in the Atlantic Forest biome within the hydrographic basins of the Almada, Cachoeira, and Una rivers in existing databases hinders the proper classification of these areas. Therefore, this study aims to analyze the classification of cabruca areas in these basins using images of different spatial resolutions and machine learning techniques. The methodology includes defining land cover and land use classes,

acquiring satellite images, processing the images using Python-based algorithms, and evaluating the supervised classification. The results show that clipping algorithms are effective in processing satellite images while preserving essential geographic information. Furthermore, the Neural Network model achieved an accuracy of 92%, regardless of the diversity of spatial resolutions. However, it is recommended that the classification results of cabruca areas be used as estimates, due to the difficulty of the algorithms in distinguishing these areas from Dense Ombrophilous Forest.

Keywords:

Artificial Intelligence, Geoprocessing, Shaded Cocoa, Biodiversity Conservation.

I. INTRODUÇÃO

A classificação das áreas de cabruca é essencial para avaliação do impacto das práticas agrícolas na conservação da biodiversidade e na manutenção dos serviços ecossistêmicos (Valadares, 2016; Embrapa, 2021; Xavier; Nascimento Jr; Chiapetti, 2021a). A cabruca, um sistema agroflorestral da região Sul baiana, representa uma estratégia multifuncional para a produção de cacau e conservação dos remanescentes florestais. A prática da cabruca, que envolve o cultivo de cacau sob a sombra da vegetação ombrófila densa, se destaca em áreas onde as florestas são escassas e fragmentadas, como uma abordagem sustentável para a agricultura e a conservação ambiental de espécies nativas (Xavier; Nascimento Jr; Chiapetti, 2021b).

Dessa forma, o geoprocessamento torna-se uma ferramenta crucial para a análise das áreas de cabruca e da Floresta Ombrófila Densa (FOD), possibilitando *insights* sobre a capacidade de sequestro de carbono e o nível de conservação de cada uma. No entanto, a classificação dessas áreas enfrenta dificuldades em identificar características que as diferenciem, devido às similaridades espectrais entre as vegetações e às limitações das técnicas de sensoriamento remoto, o que prejudica a interpretação precisa dos índices espectrais nas imagens de satélite (Valadares, 2016; Lisboa, 2022).

A dificuldade na distinção entre cabruca e FOD é identificada nos trabalhos de Souza et al. (2011), Teixeira et al. (2013), nos quais ambas as classes foram categorizadas como uma só, e em Fonseca et al. (2023), onde é recomendado seu uso como uma estimativa do estado do “cacau sombreado”. A complexidade desta distinção obriga o pesquisador à agregação das classes supracitadas. Contudo, essa abordagem compromete a precisão das análises e intervenções direcionadas à conservação e manejo desses ambientes. Têm-se as técnicas de aprendizagem de máquina e rede neurais, que apresentam vantagens quanto a precisão na classificação supervisionada de imagens de satélite (Ouchra; Belangour; Erraissi, 2023; Bhosale; Patankar, 2023).

Os investimentos em tecnologias de sensoriamento remoto e em métodos de aprendizagem de máquinas são fundamentais para o desenvolvimento de abordagens estruturadas em um viés teórico, como a codificação de resoluções espectrais diferentes em estruturas matemáticas homogêneas, garantindo que todos os elementos do *array* multidimensional sejam do mesmo tipo de dado. Neste sentido, tem-se como hipótese que a diversidade de resolução possa oferecer características espectrais distintas, contribuindo para uma análise mais detalhada. Um dos principais desafios é a baixa disponibilidade de bancos de dados dedicados, que incluam classes de cabruca e FOD, configurados de modo a conter imagens de diferentes resoluções.

Neste contexto, ao avaliar especificamente as áreas do bioma Mata Atlântica, particularmente nas Bacias Hidrográficas do Rio Almada (BHRA), do Rio Cachoeira (BHRC) e do Rio de Una (BHRU), identifica-se que os sistemas naturais da região cacaueira, como as florestas secundárias, ainda não estão rotulados, o que dificulta a classificação desses elementos. Dessa forma, o objetivo deste trabalho é analisar a classificação das áreas de cabruca e FOD nas referidas bacias, associando-as ao uso de aprendizado de máquina e a um banco de dados com diferentes resoluções espectrais.

II. MATERIAIS E MÉTODO

Para este trabalho, adota-se uma abordagem estruturada em quatro etapas lógicas. Iniciando-se com a determinação das classes de cobertura e uso da terra, bem como das áreas de coleta. Logo após, obtêm-se as imagens de satélite com diferentes resoluções. Em seguida, realiza-se o processamento das imagens. Finalizando-se com a avaliação da base de dados, e a classificação das áreas de estudo, por meio de modelo de classificação supervisionada (Figura 1).

Para a execução das etapas, utiliza-se as plataformas para pesquisa: a) Google Earth Engine, Google Colaboratory; e, b) software *Quantum Geographical Information System* (QGIS) (versão 3.22.16), licença Pública Geral (GNU).

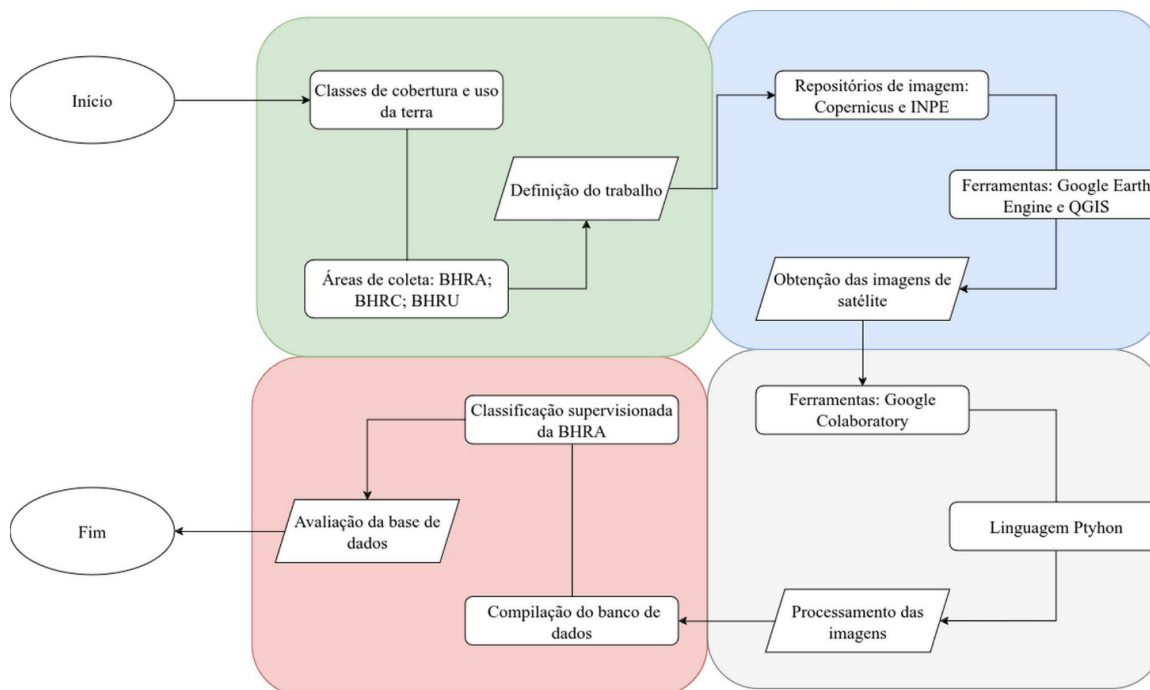


Figura 1 – Fluxograma da metodologia. Fonte: os autores 2024.

III. DEFINIÇÃO DO TRABALHO

A dinâmica das áreas de cabruca apresenta diferentes configurações, que se ajustam tanto ao proprietário da terra quanto às condições ambientais locais. Entre as configurações, destacam-se: as cabruças com remanescentes florestais da Mata Atlântica, caracterizadas por um dossel de árvores que fornece sombra para o cultivo do cacau; as cabruças raleadas, com menor adensamento devido à presença reduzida de árvores de grande porte; e as cabruças inseridas em sistemas de consórcio, onde o cacau é cultivado junto a outras espécies, como a banana (Sambuichi, 2006; Xavier; Nascimento Jr; Chiapetti, 2021a; Xavier; Nascimento Jr; Chiapetti, 2021b).

Na análise da classificação das áreas de cabruca e FOD, é fundamental compreender a relação entre essas configurações e sua reflectância. Espera-se que áreas de cabruca densa apresentem maior similaridade espectral com as áreas de FOD, enquanto as cabruças raleadas ou em consórcio apresentem características espectrais com menor similaridade. Para isso, utilizou-se um banco de dados geográficos dedicado, contendo imagens de cobertura e uso da terra da área de estudo, além do reconhecimento em campo das diferentes configurações presentes em cada imagem. As informações são extraídas das BHRA, BHRC e BHRU, sendo rotuladas conforme suas respectivas classes, seguindo o padrão estabelecido pelo Instituto Brasileiro de Geografia e Estatística (IBGE).

IV. ÁREA DE ESTUDO

A área de estudo está localizada no Sul do estado da Bahia, abrangendo unidades inseridas no bioma da Mata Atlântica (Figura 2). A BHRA engloba os municípios de Almadina, Coaraci, Itajuípe, Uruçuca, Barro Preto e Ilhéus, com mais de 247 mil habitantes. A BHRC abrange os municípios de Barro Preto, Buerarema, Firmino Alves, Floresta Azul, Ibicaraí, Ilhéus, Itabuna, Itaju da Colônia, Itapé, Itororó, Jussari, Santa Cruz da Vitória e São José da Vitória, e uma população de mais 470 mil habitantes. Já a BHRU compreende os municípios de Arataca, Santa Luzia, São José da Vitória e Una, com aproximadamente 47 mil habitantes (IBGE, 2022).

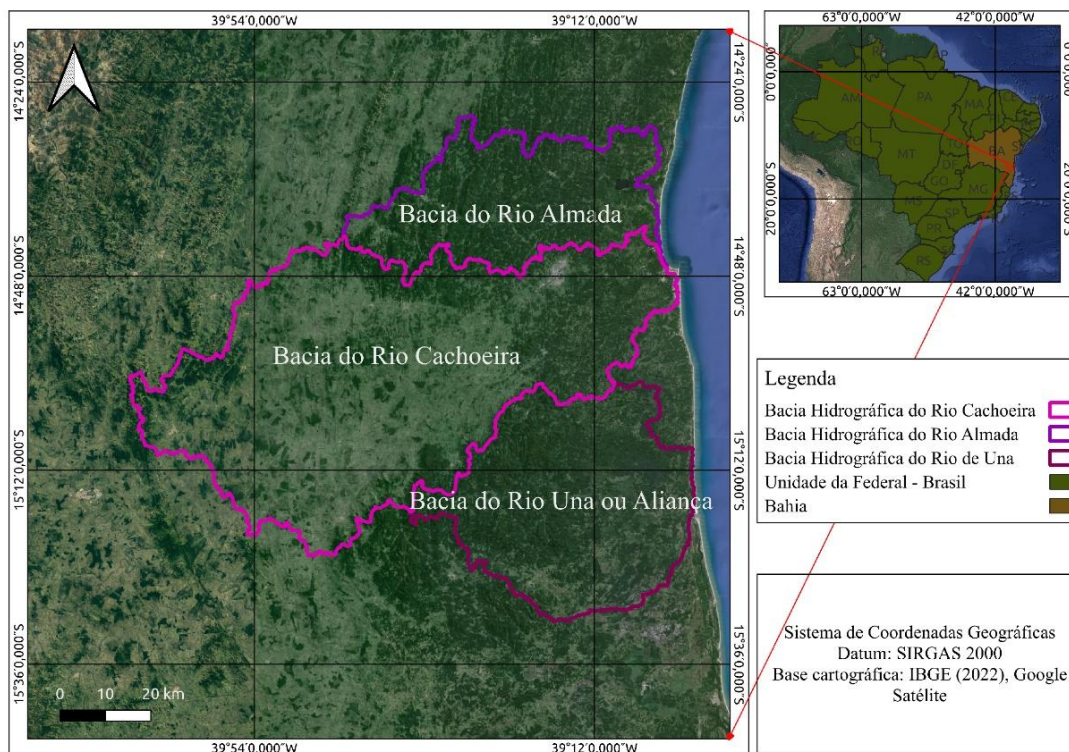


Figura 2 – Mapa de localização dos limites das bacias. Fonte: IBGE, 2022. Google Satélite, 2023. Elaboração: os autores, 2024.

V. OBTENÇÃO DAS IMAGENS DE SATÉLITES

Para a construção do *banco de dados*, foram utilizadas as imagens provenientes dos satélites CBERS04A e Sentinel-2, com resoluções de 2, 8, 10, 20, 55 e 60 metros, obtidas no INPE e Copernicus. A seleção baseou-se na hipótese que a diversidade de resolução possa oferecer características espectrais diferentes.

Na aquisição das imagens, foram avaliados o percentual de cobertura de nuvens e a localização dos pontos de interesse, garantindo a ausência de sobreposição entre as classes de infraestruturas urbanas, áreas de FOD, zonas validadas em campo e corpos d'água. As imagens com resolução de 2 metros foram geradas pelo processamento da banda pancromática (PAN) por meio do algoritmo *Pansharpening*, combinadas com informações de cor das imagens de 8 metros, obtidas pela fusão das bandas vermelha, verde e azul. Da mesma forma, obteve-se as imagens de 20 e 60 metros junto a imagem de 10 metros do Sentinel-2 (Tabela 1).

Tabela 1 – Informações das imagens utilizadas no trabalho.

Resolução	Sensor óptico	Comprimento de onda					Satélite	Data de aquisição
		Banda B	Banda G	Banda R	Banda NIR	Banda PAN		
8 metros	WPM	0,45-0,52µm	0,52-0,59µm	0,63-0,69µm	0,77-0,89µm	0,45-0,90 µm	CBERS04A	06-22-2023
8 metros	WPM	0,45-0,52µm	0,52-0,59µm	0,63-0,69µm	0,77-0,89µm	0,45-0,90 µm	CBERS04A	05-13-2021
8 metros	WPM	0,45-0,52µm	0,52-0,59µm	0,63-0,69µm	0,77-0,89µm	0,45-0,90 µm	CBERS04A	07-02-2020
10 metros	MSI	0,49µm	0,56µm	0,66µm	0,83µm	-	Sentinel-2	14-05-2024
20 metros	MSI	0,49µm	0,56µm	0,66µm	0,83µm	-	Sentinel-2	14-05-2024
60 metros	MSI	0,49µm	0,56µm	0,66µm	0,83µm	-	Sentinel-2	14-05-2024
55 metros	WFI	0,45-0,52µm	0,52-0,59µm	0,63-0,69µm	0,77-0,89µm	-	CBERS04A	13-07-2023

Fonte: INPE, 2024. ESA, 2024. Elaboração: os autores, 2024.

Além disso, para garantir a confiabilidade dos dados geográficos, as imagens foram corrigidas para refletância de superfície por meio do plugin *Semi-Automatic Classification Plugin* (SCP). As coordenadas das áreas de cabruca foram coletadas durante as campanhas de validação em campo, utilizando o aplicativo *Timestamp Camera Free* e o GPS portátil *eTrex 10*. Posteriormente, os pontos foram conferidos no *Google Earth Pro*, a exemplo das coordenadas -14°47'43,536''S -39°10'21,24''O e -14°37'42,072''S -39°32'21,056''O.

VI. PROCESSAMENTO DAS IMAGENS

As áreas de cabruca e FOD identificadas durante as visitas de campo são posteriormente mapeadas utilizando as ferramentas do QGIS. Em seguida, para reduzir a probabilidade de erro humano, devolveu-se um algoritmo na linguagem Python para realizar a leitura e o recorte automático das imagens correspondentes, ajustando-as para serem usadas no treinamento dos modelos de aprendizado de máquina. Esse algoritmo permite a definição manual das dimensões das imagens, conferindo flexibilidade e controle ao processamento. Além disso, organiza de modo eficiente o conjunto de dados ao armazenar as imagens recortadas em pastas separadas conforme suas respectivas classes.

Ressalta-se que, além das classes principais, o banco de dados também inclui classes de infraestrutura urbana, pastagem e corpos d'água presentes nas bacias. A delimitação dessas áreas baseia-se em visitas de

campo, conhecimento prévio do local e informações visuais da Coleção 9 do projeto MapBiomas. Além disso, a classe "nuvem" foi incorporada devido à alta frequência desse fenômeno nas imagens de satélite, influenciada pelo clima tropical úmido da região Sul da Bahia e pela umidade proveniente do Oceano Atlântico.

VII. AVALIAÇÃO DA BASE DE DADOS

Após a compilação dos dados, as informações dos *pixels* das imagens são acessadas por meio da leitura das bandas espectrais e armazenadas em uma variável do tipo *NumPy*, um *array* multidimensional amplamente utilizado em Python para a manipulação eficiente de dados numéricos. Essa estrutura é essencial para tarefas de processamento de imagens e aprendizado de máquina, pois permite a realização de operações matemáticas e lógicas de forma otimizada, como a manipulação de grandes volumes de dados espaciais. A partir desse ponto, são aplicados algoritmos de aprendizado de máquina para avaliar a eficácia do banco de dados construído. Utilizam-se modelos como *K-nearest neighbors* (KNN), que baseiam suas previsões na similaridade entre pontos de dados vizinhos, e Redes Neurais, capazes de identificar padrões mais complexos e sutis nos dados (Alba et al., 2022; Fagundes; Júnior, 2022).

Para garantir o melhor desempenho e otimizar a precisão na classificação, os hiperparâmetros dos modelos foram ajustados conforme as diretrizes da biblioteca *scikit-learn*, que oferece ferramentas robustas para a otimização de modelos de aprendizado de máquina. Entre os métodos utilizados, destaca-se a validação cruzada, que assegura uma avaliação robusta do modelo, e o ajuste de hiperparâmetros por meio da busca em grade (*Grid Search*), que permite encontrar a combinação ideal de parâmetros para maximizar a acurácia no reconhecimento de padrões nas imagens de satélite, evitando *overfitting* (Homer et al., 2019; Smith; Jones, 2021).

O conjunto de dados foi dividido em três partes: 70% para treinamento, 15% para validação e 15% para teste, garantindo um balanceamento adequado entre as classes. A avaliação do desempenho dos modelos foi realizada com base em métricas como *acurácia*, *precisão*, *recall* e *F1-score*, permitindo uma comparação detalhada da eficácia dos modelos na classificação das imagens de satélite.

Esses algoritmos são empregados para identificar padrões de *pixels* nas imagens com resolução de 8 e 10 metros. Na classificação das imagens, avalia-se duas condições distintas: a condição A, que corresponde à região próxima ao município de Itabuna-BA, caracterizada principalmente pela presença predominante das

classes FOD, infraestrutura urbana e pastagem, com pouca ocorrência de áreas de cabruca; e a condição B, que se refere a área próxima da CEPLAC-BA, composta por FOD, cabruca e infraestrutura urbana.

Além disso, para uma segunda rodada de avaliações, considerando as correções das variáveis estruturais identificadas e questionando a hipótese da diversidade de resoluções, utiliza-se um banco de dados composto por imagens de 2 metros de resolução, obtidas de um único satélite.

O desempenho é avaliado por meio das métricas de avaliação (*acurácia, precisão, recall e F1-score*) e da matriz de confusão, contribuindo para determinar o papel do banco de imagens e sua influência nos modelos de classificação com aprendizado de máquina. Por fim, realiza-se a estimativa manual da área de cada classe por meio da sobreposição dos dados obtidos em visitas de campo e da coleção 9 do MapBiomas, devido à dificuldade na obtenção de uma classificação confiável para a área de estudo. Nesta abordagem, considerou-se que 0,05% da classe FOD corresponde à área de cabruca, com base nas validações dos dados em campo.

VIII. RESULTADOS E DISCUSSÃO

Na análise dos resultados da classificação pelos modelos de aprendizado, identifica-se que os referidos algoritmos apresentam dificuldades em distinguir com precisão as áreas de cabruca. Para o modelo KNN, avalia-se que sua performance nos dados de teste alcançou uma acurácia global de 82%, o que indica a sua capacidade de fornecer estimativas próximas aos valores de referência.

Contudo, evidencia-se um desempenho inferior, abaixo de 50%, para a classe cabruca quando comparada com as demais, onde destaca-se a precisão e a disparidade reforçada no F1-Score, que revela um desequilíbrio na classificação da classe supracitada. Por outro lado, as classes de FOD, pastagem e corpo d'água, demonstram a eficácia do banco de dados com precisão superior a 80% (Tabela 2).

Tabela 2 – Resultados da avaliação do modelo de KNN com as imagens do banco de dados dedicado.

Classe	Precisão / %	Recall / %	F1-Score / %
Cabruca	48	36	41
Floresta ombrófila densa	86	84	85
Infraestrutura urbana	60	78	68
Pastagem	87	84	86
Corpo d'água	89	90	90
Nuvem	72	51	60

Autoria: os autores, 2024.

Na Tabela 2 identifica-se o desempenho do modelo em relação ao *recall* das classes cabruca e nuvem, que identificaram, respectivamente, 36% e 51% das instâncias relevantes. A baixa capacidade do modelo em

recuperar informações compromete a qualidade das classificações, especialmente para a classe cabruca, possivelmente devido à sobreposição com outras classes. A dificuldade em reivindicar características espectrais adequadas nas imagens limita ainda o processo classificatório, especialmente quando há similaridade do comportamento espectral entre diferentes classes de cobertura do solo, que não permite a separação das classes apenas com base nos valores de reflectância, exigindo que outros critérios sejam considerados, além dos espectrais, para a separação, como a fotointerpretação (Moreira; Adami; Rudorff, 2004; Sena Souza et al., 2016).

Segundo Tan et al. (2022), um dos fatores associados a esse problema é a ausência de sensores capazes de reduzir os efeitos da sobreposição de superfícies. Na classificação das imagens com resoluções de 8 e 10 metros, analisa-se que, para a condição A, onde existe pouca ocorrência da área de cabruca, surgiram algumas inconsistências, segundo a validação dos dados em campo. Na imagem de 8 metros, são analisadas discrepâncias nas classificações da classe corpo d'água, interpretadas erroneamente como classe de infraestrutura urbana, enquanto porções de áreas de pastagem são reconhecidas como corpo d'água.

Os resultados exemplificam a importância do índice de separabilidade das classes, fator relacionado com a precisão da classificação. Para Rocha et al. (2012) e Das e Pandeys (2019), a baixa precisão na classificação de pixels mistos estreitamente associados é devido à menor diferença espectral em relação às categorias classificadas corretamente, uma vez que, em ambientes heterogêneos como os urbanos, a presença de uma mistura complexa de respostas espectrais dentro de um único *pixel* (mixels) resulta numa assinatura espectral que é uma combinação dos materiais básicos de uso e ocupação do solo. Esta mistura espectral dificulta a identificação de classes por meio de técnicas analíticas pixel-a-pixel, por isso, ao abordar as classificações incorretas, considera-se o emprego dos algoritmos de dilatação morfológica, mitigando problemas de mistura de *pixels*, e utilização de técnicas para localização e remoção das sombras (Rocha et al. 2012; Liang; Liang; Sun, 2023).

Quando a resposta espectral desses *pixels mistos* é muito similar entre si ou a outras classes (puras ou mistas), a separabilidade espectral das categorias diminui, levando a confusões na classificação. Por isso, avalia-se o desempenho do modelo na estimativa das áreas de infraestrutura urbana e pastagem. No entanto, destaca-se os pontos de cabruca, próximos a classe FOD, reforçando a não identificação de padrões distintos, em consequência tem-se a sobreposição delas.

Para a imagem de 10 metros, constata-se aptidão na estimativa das áreas de infraestrutura urbana e pastagem, identificando ruas e pastos com maior qualidade em comparação com as imagens de 8 metros (Figura 3).

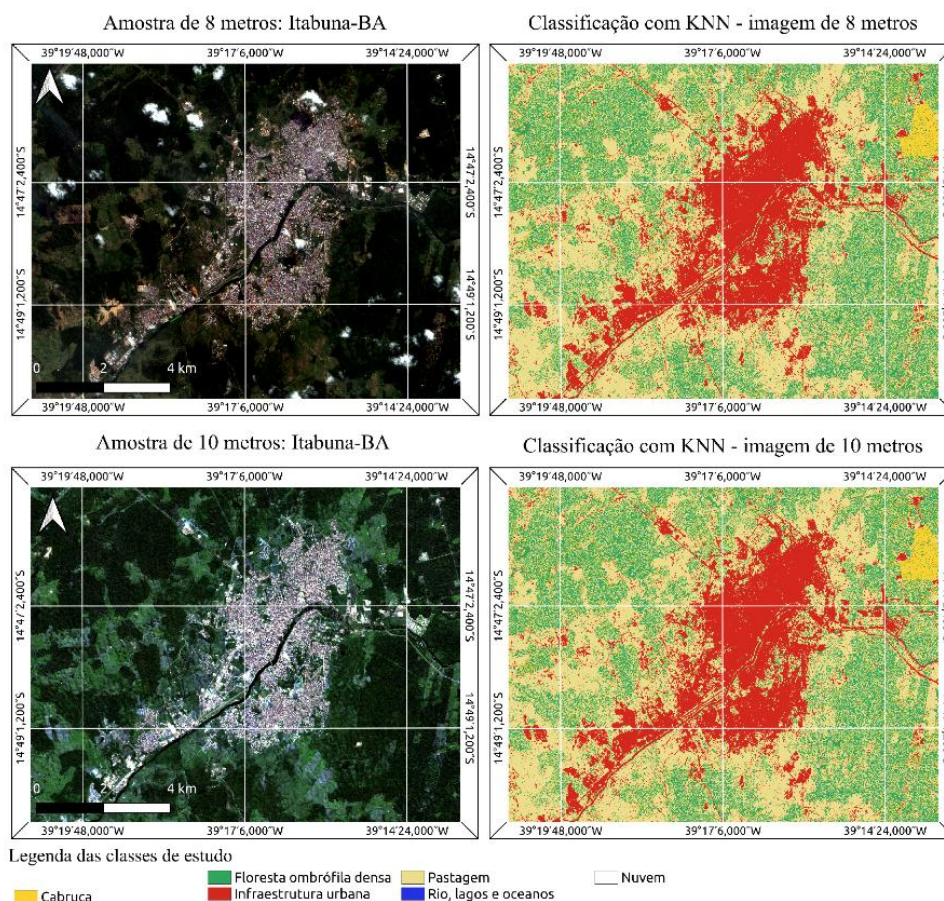


Figura 3 – Classificação da cobertura e uso da terra pelo modelo KNN. Fonte: os autores, 2024.

Para a condição B, analisa-se a capacidade de classificação das áreas de cabruca presente na imagem, alcançando o polígono definido manualmente, com base na validação dos dados em campo. Isso assinala que o modelo recuperou informações preservadas no banco de imagens, especialmente na imagem de 10 metros de resolução, que se apresenta uma predominância da classe FOD no entorno do polígono (Figura 4).

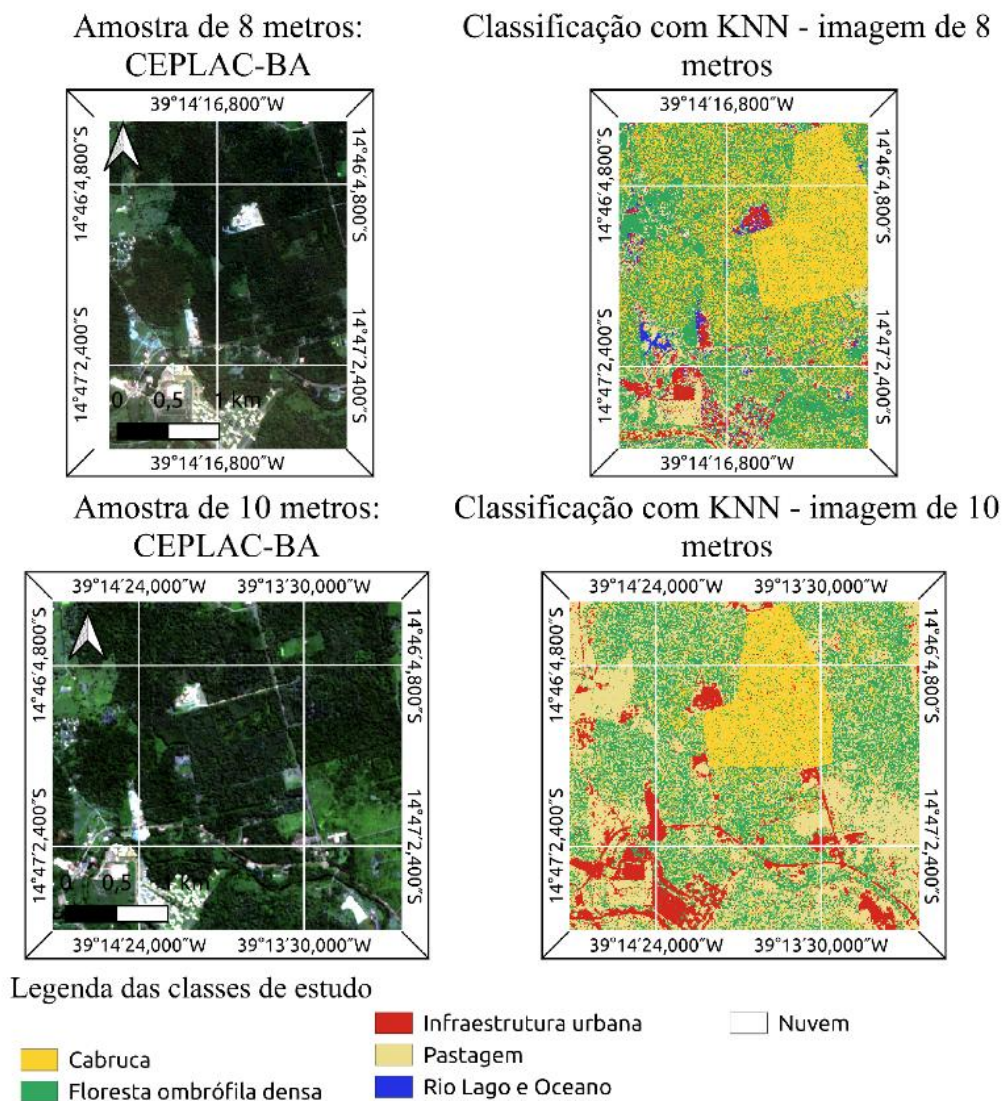


Figura 4 – Classificação da cobertura e uso da terra da área próxima à CEPLAC-BA pelo modelo de KNN. Fonte: os autores, 2024.

Para o modelo de Rede Neural, os resultados demonstram que o algoritmo é eficaz na classificação das diferentes classes, apresentando desempenho consistente e equilibrado em todas as categorias. Isso indica que a Rede Neural, em comparação com o método KNN, é uma escolha vantajosa para a classificação das áreas de cabruca e FOD. No entanto, é necessário avaliar a estrutura da rede neural; o aumento do número de camadas ocultas pode não melhorar significativamente a acurácia, gerando custos computacionais sem ganhos substanciais (Moraitis et al., 2021).

Nos dados de teste, a acurácia alcançou 92%, com uma precisão de 84% para a classe cabruca, embora o *recall* sugira a perda de informações, diferente da classe FOD, que apresenta um F1-Score equilibrado (Tabela 3).

Tabela 3 – Resultados da avaliação do modelo de Rede Neural utilizado o banco de dados

Classe	Precisão / %	Recall / %	F1-Score / %
Cabruca	84	65	73
Floresta ombrófila densa	94	92	93
Infraestrutura urbana	69	93	79
Pastagem	96	91	94
Rio, lagos e oceanos	97	95	96
Nuvem	95	86	90

Fonte: os autores, 2024

Na classificação da condição A, avalia-se que as imagens de 8 metros e 10 metros exibem características similares ao modelo KNN, com resultados inconsistentes na classificação das classes de corpo d'água e de nuvem, especialmente nas imagens de 10 metros. Identifica-se que as imagens selecionadas para compor o banco de dados não atenderam aos requisitos necessários para evitar sobreposições com outras categorias.

Salienta-se, a importância da avaliação prévia das imagens antes de serem submetidas ao algoritmo de recorte. Torna-se indispensável o estabelecimento de critérios relacionados ao nível de rugosidade ou a presença de elementos de outra classe nas imagens. Patel, Chatterjee e Gorai (2017) Nazmuzaman e Anwar (2019), enfatizam a textura, intensidade, condições de iluminação e sombras. A exemplo do rio Cachoeira, próximo ao município de Itabuna-BA, que apresenta sobreposição de nuvens e proximidade com infraestrutura urbana e FOD, o que impacta negativamente a classificação (Figura 5).

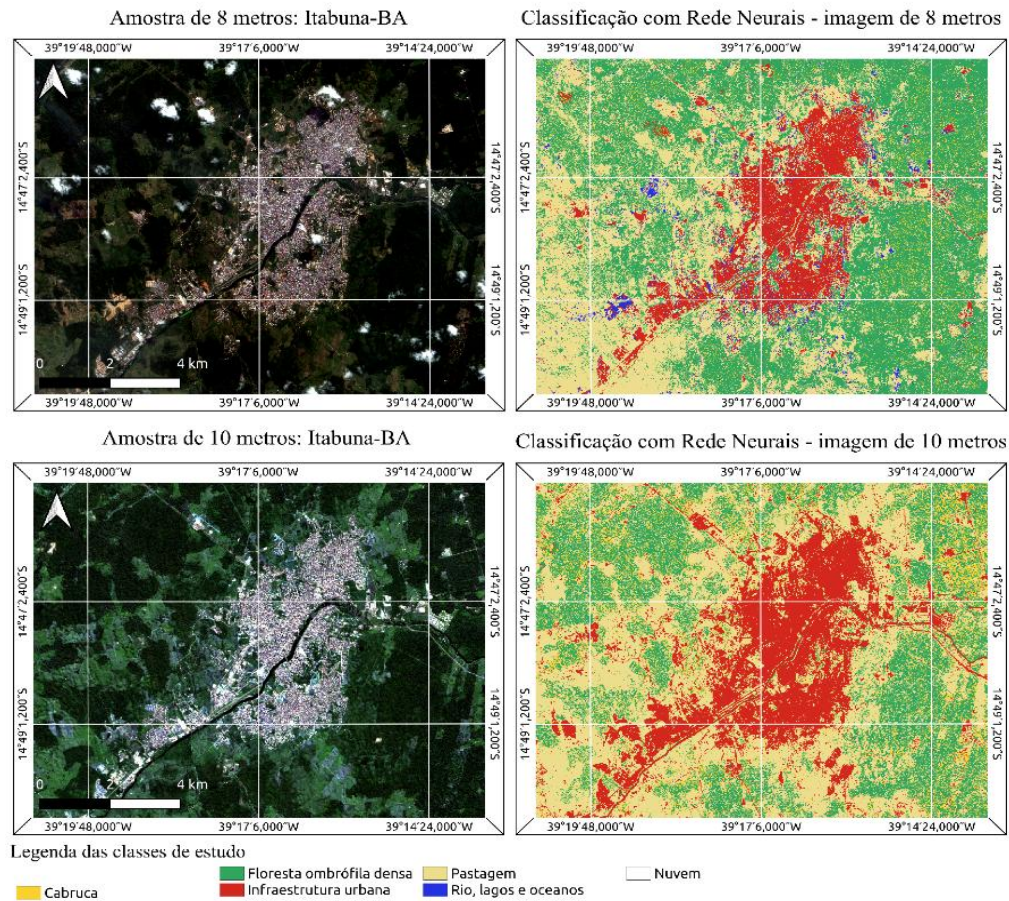


Figura 5 – Classificação da cobertura e uso da terra próxima ao município de Itabuna-BA pelo modelo Rede Neural. Fonte: os autores, 2024.

Além disso, identificam-se classificações em comum para ambos os algoritmos, especialmente na condição B, onde a mesma área de cabruca é representada por um agrupamento de *pixels*.

Ressalta-se que, a disposição dos *pixels* na imagem de 8 metros se assemelha a uma área com pouca presença de FOD, enquanto a classificação na imagem de 10 metros é mais conservadora, com *pixels* representando áreas de cabruca e FOD. Isso é corroborado com a validação dos dados em campo, onde existem áreas de cabruca e outras áreas com floresta plantadas (Figura 6).

Da mesma forma, destaca-se a qualidade na estimativa das áreas de infraestrutura urbana e pastagem, com os *pixels* definidos de forma consistente e precisa para as amostras avaliadas. Essa consistência é próxima do que foi verificado tanto em validação dos dados em campo, quanto com a utilização do Google Earth Pro.

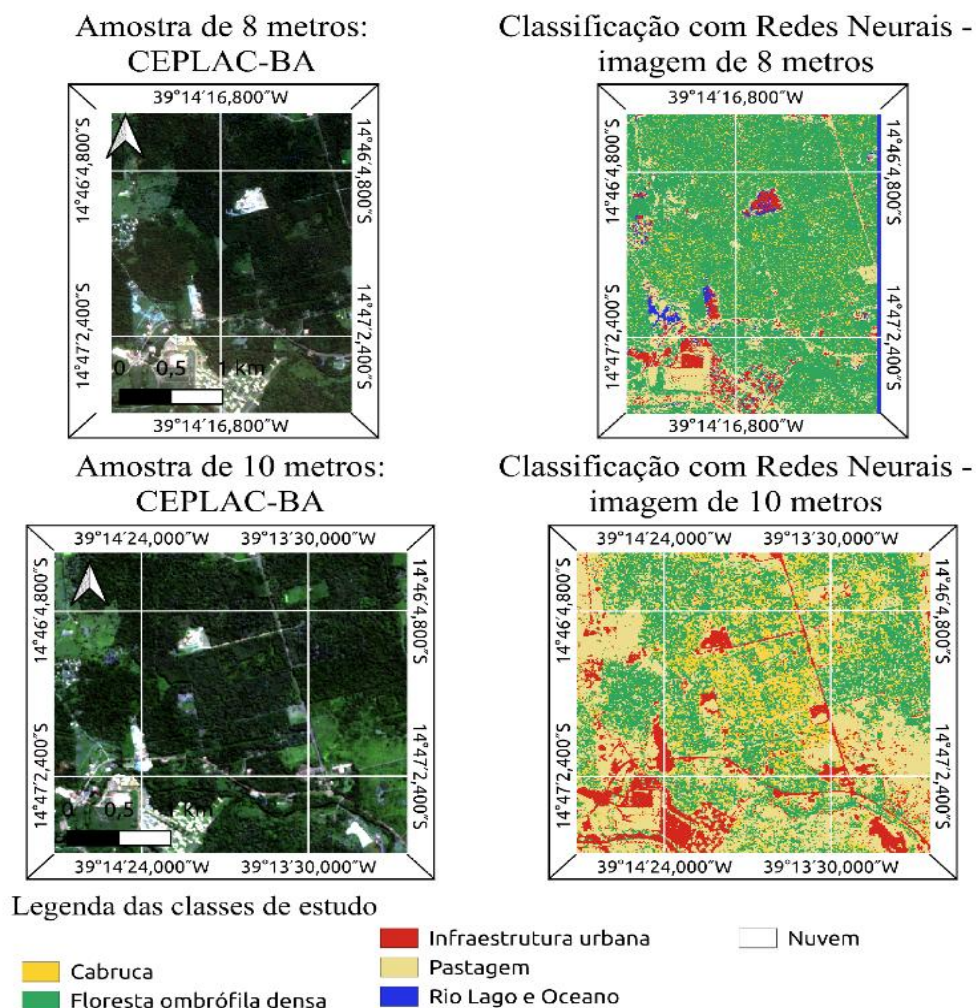


Figura 6 – Classificação da cobertura e uso da terra na área próxima à CEPLAC-BA pelo modelo Rede Neural. Fonte: Autoria: os autores, 2024.

Os resultados dos modelos explicitam a capacidade da classificação das imagens de resoluções distintas. Nas imagens de 8 metros de resolução avalia-se a incapacidade do modelo no reconhecimento das vizinhanças mais próximas, em detrimento da maior densidade de *pixels*. Além disso, os detalhes incompatíveis entre as imagens de 2 e 8 metros com as de 55 e 60 metros de resolução interferem na análise precisa da classificação das imagens.

Desta forma, na segunda rodada de avaliação, identifica-se uma acurácia de 86% para o algoritmo de Redes Neurais com as imagens de 2 metros de resolução. Analisa-se o equilíbrio na classificação de todas as classes, exceto a classe cabruca, cuja precisão foi similar à do modelo KNN. Além disso, ressalta-se, que, para manter a hipótese de que a diversidade de resoluções oferece características espectrais complementares que enriquecem a análise, não foram aplicadas técnicas de reamostragem.

Neste sentido, ressalta-se os preditores empregados, destacam-se as variáveis de textura, rugosidade e densidade de *pixels*, que mostraram desempenho inferior na classificação de classe cabruca quando utilizadas de forma homogênea, sem a combinação com informações provenientes de imagens de outras resoluções. Destaca-se, assim, a precisão obtida no modelo de Redes Neurais que utilizou diferentes resoluções para o treinamento.

Nesse contexto, destaca-se que os preditores utilizados, como textura, rugosidade e densidade de pixels, apresentaram baixo desempenho na classificação da classe *cabruca* quando aplicados de forma homogênea, sem a combinação de informações provenientes de imagens com outras resoluções. Isso evidencia a maior acurácia alcançada pelo modelo de Rede Neural treinado com múltiplas resoluções (Tabela 4).

Tabela 4 – Resultados da avaliação do modelo de Redes Neurais com as imagens de 2 metros de resolução

Classe	Precisão / %	Recall / %	F1-score / %
Cabruca	49	21	30
Floresta ombrófila densa	85	90	88
Infraestrutura urbana	88	84	86
Pastagem	89	88	88
Corpo d'água	82	93	87
Nuvem	91	91	91

Autoria: os autores, 2024.

Na Tabela 4, avalia-se que a classe pastagem exibe métricas robustas, para os valores de precisão, *recall* e F1-score, espelham um desempenho consistente do modelo na identificação dessa categoria. Para as demais classes, salienta-se que, além das variáveis espaciais que afetam a qualidade da classificação, é relevante a consideração das características comuns compartilhadas entre elas.

Na aplicação do modelo, analisam-se classificações corretas para as classes de corpo d'água e nuvem, devido à aplicação dos ajustes identificados nas avaliações anteriores. Esses ajustes incluíram o estabelecimento de critérios que realçaram as características das classes. Segundo Condé et al. (2023), tais como cor, tonalidade ou intensidade de brilho, refletância, textura, tamanho, forma e influência do entorno.

Assim, identifica-se a drenagem do rio Cachoeira e uma classificação conservadora, com as áreas de cabruca associadas a presença da classe FOD (Figura 7).

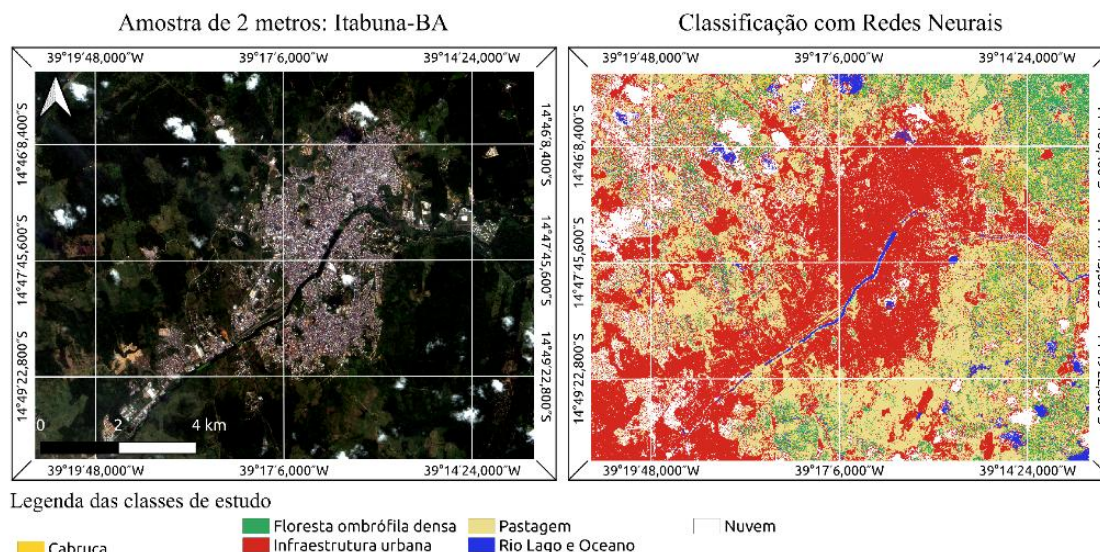


Figura 7 – Classificação da cobertura e uso da terra pelo modelo Redes Neurais com resolução de 2 metros. Fonte: os autores, 2024.

Na análise da matriz de confusão (Figura 8), a classe cabruca é sobreposta às outras classes. Li, Ustin e Lay (2005), Valadares (2016) e Shivakumar e Rajashekararadhya (2017), acentuam que as similaridades espectrais entre vegetações influenciam as limitações da classificação e que a probabilidade da produção de classificações incorretas é maior para classes espectrais dependentes. Uma das condições que reforça tal suposição é devido às árvores utilizadas como sombreamento apresentarem copas que dificultam a obtenção de características próprias da classe cabruca.

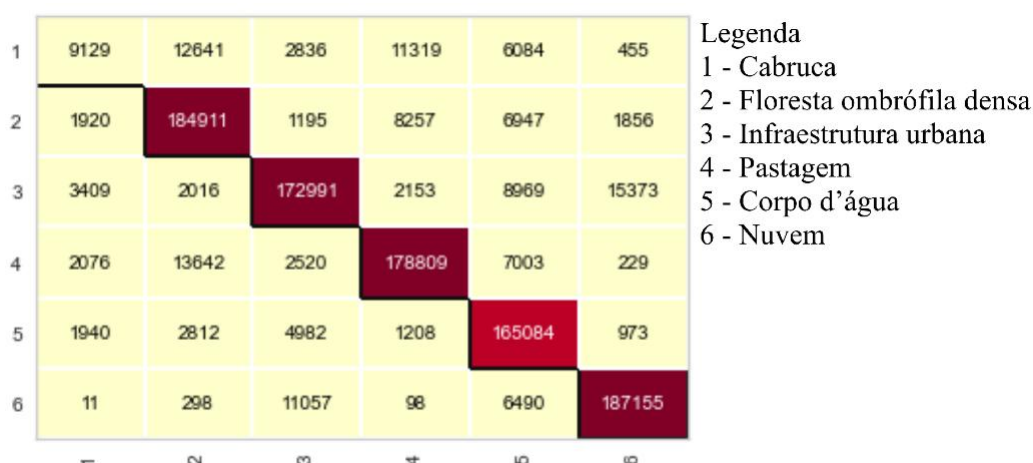


Figura 8 – Matriz de confusão para avaliação do desempenho de modelos de Redes Neurais, utilizando imagens com resolução espacial de 2 metros. Autoria: os autores, 2024.

Além disso, para lidar com a interferência de iluminação desequilibrada e ruído, Shi e Pun (2018) propõem uma nova função de perda percentual baseada em similaridade espectral e estrutural, enquanto Liu

et al. (2024) apresentam três novos índices para uma extração mais precisa da vegetação. Tais abordagens possibilitam a correção desses parâmetros nas imagens utilizadas para o treinamento. No entanto, devido à forte similaridade espectral entre cabruca e FOD, diferenciá-las permanece como o principal desafio. Conforme Reuss-Strenzel e Faria (2023), as respostas espectrais da cabruca são semelhantes às respostas espectrais de fragmentos florestais da Mata Atlântica.

Na sumarização dos resultados, avalia-se que a classificação do modelo de Redes Neurais, na imagem de 8 metros, aproxima-se da área estimada para cada classe. A estimativa manual apresenta valores com base em interpretação visual, o que deixa a desejar para a classe de cabruca. A consideração de que 0,05% da classe FOD corresponde à classe cabruca ajuda na parametrização de um valor fixo para a avaliação dos algoritmos, auxiliando em ajustes para melhorar a precisão da classificação e na estimativa do erro (Figura 9).

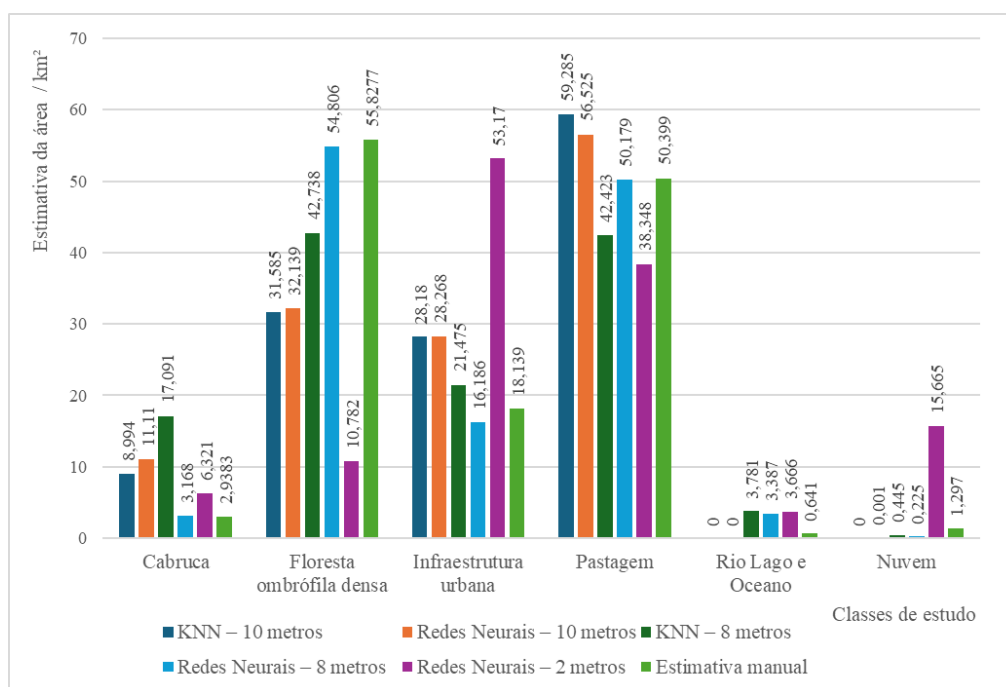


Figura 9 – Comparação entre as classificações da cobertura e uso da terra pelos modelos de aprendizagem de máquina. Fonte: os autores, 2024.

Na Figura 9, analisa-se que os erros percentuais da classe FOD e de pastagem foram respectivamente, de 1,83% e 0,44% menor do que o valor de referência, evidenciando a superioridade do modelo em comparação com os demais. Em contraste, a classe cabruca com um erro percentual de 7,82%, assinala a necessidade de correções quanto a classificação dessa classe.

Identifica-se a sinergia dos resultados com Fonseca et al. (2023), na classificação das áreas com cacau. Salienta-se o termo “cacau sombreado” que inclui tanto áreas de cacau cabruca quanto de consórcio. No

entanto, essa abordagem pode dificultar intervenções direcionadas à conservação e manejo dos ambientes com cacau cabruca.

Além disso, avalia-se a limitação na abrangência das BHRA, BHRC e BHRU devido a extensão aproximada de 764 mil km² e a necessidade de processamento computacional robusto para armazenamento dos *pixels* na memória. Por isso, Mielikainen (2006), ressalta-se estratégias para armazenar os valores reais de pixel por meio dos esquemas de banda intercalada por linha, banda intercalada por pixel ou banda sequencial.

IX. CONCLUSÕES

Os resultados demonstram a eficácia do uso do algoritmo de recorte para o desenvolvimento do banco de dados geográfico, adaptando-se às características específicas da área de estudo. Destaca-se a importância de estabelecer critérios que realcem as características e atributos de cada classe, auxiliando no reconhecimento de padrões que aprimoram o desempenho da classificação.

A classificação das áreas de infraestrutura urbana e pastagem evidencia o potencial do banco de dados como uma ferramenta eficaz para o treinamento de modelos de aprendizado de máquina. Contudo, observa-se dificuldade em distinguir com precisão a classe cabruca, como mostrado nas tabelas, tanto para os algoritmos KNN quanto para Redes Neurais com imagens de 2 metros de resolução. Isso reforça que os resultados devem ser interpretados como estimativas da presença de cabruca.

Por outro lado, para as demais classes, como infraestrutura urbana, pastagem e FOD, avalia-se que o modelo de Redes Neurais apresentou o melhor desempenho, independentemente da diversidade de resoluções. Adicionalmente, recomenda-se o uso de imagens de 10 metros de resolução para a classificação dessas classes, em vez da densidade de pixels das imagens de 8 metros, que demonstram a incapacidade do modelo de determinar com precisão padrões ou vizinhanças na classificação do uso e cobertura do solo.

Além disso, sugere-se adotar estratégias eficazes para a leitura e armazenamento das informações na memória do computador ou em servidores, como métodos alternativos para leitura de informações das imagens, especialmente para áreas extensas que exigem recursos computacionais robustos, como exemplificado pela área de estudo de 8.600 km². Por fim, ressalta-se a importância da colaboração interdisciplinar e da sinergia com outros trabalhos que abordem outras áreas do bioma Mata Atlântica, como manguezais e restingas

Agradecimentos

Ao Conselho Nacional de Desenvolvimento Científico e Tecnológico (CNPQ) pela bolsa de produtividade para um dos autores. À Coordenação de Aperfeiçoamento de Pessoal de Nível Superior (CAPES) pela concessão de bolsa de pós-graduação, para um dos autores. À Fundação de Amparo de Pesquisa da Bahia (FAPESB) pela bolsa de mestrado a um dos autores.

X. REFERÊNCIAS

- ALBA, E.; ALEXANDRE, M. L. S.; MARCHESAN, J.; SOUZA, L. S. B.; BEZERRA, A. C.; SILVA, E. A. Comparação en-tre Algoritmos de Aprendizado de Máquina para a Identificação de Floresta Tropical Sazonalmente Seca. *Anu-ário do Instituto de Geociências*, vol. 45, no. 1982-3908, p. 1–10, 2022. DOI https://doi.org/10.11137/1982-3908_2022_45_4075.
- BHOSALE, V.; PATANKAR, A. Performance Comparision of VGG-16 and ResNet-34 Algorithms for Supervi-sed Classification of Landsat Images. 16 Jul. 2023. DOI <https://doi.org/10.1109/igarss52108.2023.10281537>.
- CHATTERJEE, S.; BHATTACHARYYA, S.; SAHA, S.; HALDER, A.; RANJAN, P. Detection of Dense Built-Up Area in Low-Resolution Satellite Images Using Deep Learning and DBSCAN Approaches. *International Journal of Pat-tern Recognition and Artificial Intelligence*, vol. 38, no. 04, 30 Mar. 2024. <https://doi.org/10.1142/s0218001424510042>.
- CONDÉ, T. M.; HIGUCHI, N.; ADRIANO; ASSIS, A.; JACKELIN DIAS CONDÉ; CAMARGO, A.; CARNEIRO, L. Spectral Patterns of Pixels and Objects of the Forest Phytophysiognomies in the Anauá National Forest, Roraima State, Brazil. *Ecologies*, vol. 4, no. 4, p. 686–703, 28 Oct. 2023. DOI <https://doi.org/10.3390/ecologies4040045>.
- DAS, P.; PANDEY, V. Use of Logistic Regression in Land-Cover Classification with Moderate-Resolution Multis-pectral Data. *Journal of the Indian Society of Remote Sensing*, vol. 47, no. 8, p. 1443–1454, 8 May 2019. <https://doi.org/10.1007/s12524-019-00986-8>.
- FAGUNDES, W. S.; JÚNIOR, M. J. A. Evolução das Técnicas de Classificação de Imagens do Sensoriamento Re-moto Utilizadas na Produção Científica Brasileira: Evolution of remote sensing image classification techniques used in Brazilian scientific production. *Geosciences = Geociências*, vol. 41, no. 3, p. 593–604, 2022. DOI <https://doi.org/10.5016/geociencias.v41i03.16209>.
- FONSECA, M. G. et al. Mapbiomas Cacau: avanços e desafios no mapeamento do cultivo sombreado de cacau no Sul da Bahia. [S.l.]: [s.n.], 2023.
- GAIAD, S.; CARVALHO, P. E. R. Glossário — Portal Embrapa. 22 dez. 2021. Disponível em: <https://www.embrapa.br/>.
- GOOGLE COLABORATORY. Disponível em: <https://colab.research.google.com/>. Acesso em: 25 out. 2025.
- GOOGLE EARTH ENGINE TEAM. Google Earth Engine: a planetary-scale geospatial analysis platform. [S.l.]: [s.n.], 2013. Disponível em: <https://earthengine.google.com/>. Acesso em: 25 out. 2025.
- HAFSA, O.; ABDESSAMAD, A. E. B. Machine learning algorithms for satellite image classification using Google Earth Engine and Landsat satellite data: Morocco case study. *IEEE Access*, v. 11, p. 71127–71142, 2023. <https://doi.org/10.1109/ACCESS.2023.3293828>.

LI, L.; USTIN, S. L.; LAY, M. Application of multiple endmember spectral mixture analysis (MESMA) to AVIRIS imagery for coastal salt marsh mapping: a case study in China Camp, CA, USA. *International Journal of Remote Sensing*, vol. 26, no. 23, p. 5193–5207, 10 Dec. 2005. <https://doi.org/10.1080/01431160500218911>.

LIANG, J.; LIANG, G.; SUN, L. Using Sentinel images for analyzing water and land separability in an agricultural river basin. *Environmental Monitoring and Assessment*, vol. 195, no. 11, 13 Oct. 2023. <https://doi.org/10.1007/s10661-023-11908-0>.

LISBOA, G. P. Validação do Modelo Linear de Mistura Espectral e Índices De Vegetação para Identificação de Fragmentos Florestais de Mata Atlântica e Sistema Agroflorestal Cacau Cabruca Utilizando Imagens Orbitais. 2022.

LIU, X.; ZHENG, X.; WANG, Z.; LI, Z.; WANG, K.; ZHANG, H.; DUAN, J. Monitoring Wetland Changes and Analyzing the *Spartina alterniflora* Invasion in the Yellow River Delta over the Past 30 Years Based on Google Earth Engine. *IEEE Geoscience and Remote Sensing Letters*, vol. 21, p. 1–5, 1 Jan. 2024. DOI <https://doi.org/10.1109/lgrs.2024.3400027>.

MIELIKAINEN, J. Lossless compression of hyperspectral images using lookup tables. *IEEE Signal Processing Letters*, vol. 13, no. 3, p. 157–160, Mar. 2006. <https://doi.org/10.1109/lsp.2005.862604>.

MORAITIS, N.; TSIPI, L.; VOUYIOUKAS, D.; GKIONI, A.; LOUVROS, S. Performance evaluation of machine learning methods for path loss prediction in rural environment at 3.7 GHz. *Wireless Networks*, vol. 27, no. 6, p. 4169–4188, 18 Jun. 2021. <https://doi.org/10.1007/s11276-021-02682-3>.

MOREIRA, M. A.; ADAMI, M.; RUDORFF, B. F. T. Análise espectral e temporal da cultura do café em imagens Landsat. *Pesquisa Agropecuária Brasileira*, v. 39, n. 3, p. 223–231, mar. 2004. <https://doi.org/10.1590/S0100-204X2004000300004>.

NAZMUZZAMAN, K. M.; ANWAR, S. Robust Weed Recognition Through Color Based Image Segmentation and Convolution Neural Network Based Classification. Nov. 2019.

PATEL, A. K.; CHATTERJEE, S.; GORAI, A. K. Development of machine vision-based ore classification model using support vector machine (SVM) algorithm. *Arabian Journal of Geosciences*, vol. 10, no. 5, Mar. 2017. <https://doi.org/10.1007/s12517-017-2909-0>.

ROCHA, J.; TENEDÓRIO, J. A.; ESTANQUEIRO, R.; SOUSA, P. M. Classificação de uso do solo urbano através da análise linear de mistura espectral com imagens de satélite. *Finisterra*, v. 42, n. 83, 13 dez. 2012. <https://doi.org/10.18055/finis1438>.

SAMBUICHI, R. H. R. Estrutura e dinâmica do componente arbóreo em área de cabruca na região cacauzeira do sul da Bahia, Brasil. *Acta Botanica Brasilica*, vol. 20, no. 4, p. 943–954, Dec. 2006. <https://doi.org/10.1590/s0102-33062006000400018>.

SENA SOUZA, J. P. et al. Comparação dos métodos de classificação por ângulo espectral e distância euclidiana no mapeamento das formas de terreno. *Revista Brasileira de Geomorfologia*, v. 17, n. 3, 1 set. 2016. <https://doi.org/10.20502/rbg.v17i3.846>.

SHI, C.; PUN, C.-M. Perceptual Loss for Superpixel-Level Multispectral and Panchromatic Image Classification. 1 Apr. 2018. DOI <https://doi.org/10.1109/icassp.2018.8461966>. Available at: <https://ieeexplore.ieee.org/document/8461966>.

SHIVAKUMAR, B. R.; RAJASHEKARARADHYA, S. V. Spectral similarity for evaluating classification performance of traditional classifiers. 1 Mar. 2017. DOI <https://doi.org/10.1109/wispnet.2017.8300111>.

SMITH, A. R.; JONES, M. L. Optimization techniques in machine learning for satellite image analysis. Computational Intelligence in Remote Sensing, vol. 8, p. 125–138, 2021. <https://doi.org/10.1016/j.cirs.2021.06.004>.

SOUZA, C. M. P. de; MOREAU, M. S.; MOREAU, A. M. S. dos S.; FONTES, E. de O. Níveis de degradação de pastagens da bacia do Rio Colônia-BA com uso de imagens LANDSAT 5 TM. Revista Brasileira de Geografia Física, v. 3, n. 3, p. 228–243, 7 jul. 2011. <https://doi.org/10.26848/rbgf.v3i3.232685>.

TAN, Z.; LIU, C.; MA, S.; WANG, X.; SHANG, J.; WANG, J.; AI, W.; YAN, W. Detecting Multilayer Clouds From the Geostationary Advanced Himawari Imager Using Machine Learning Techniques. IEEE Transactions on Geoscience and Remote Sensing, vol. 60, p. 1–12, 2022. DOI <https://doi.org/10.1109/TGRS.2021.3087714>.

TEIXEIRA, A. C. de O. et al. Análise da dinâmica de uso e ocupação da terra na bacia hidrográfica do baixo Rio de Contas – Bahia, entre os anos de 1973 a 2001. Revista Geoaraguaia, 18 set. 2013.

VALADARES, J. O. Método de classificação especialista para imagens de sensoriamento remoto para o mapeamento de áreas de cultivo de cacau-cabruca. repositorio.ufba.br, 20 Jun. 2016.

XAVIER, L. B.; NASCIMENTO JR, F. D. C. do; CHIAPETTI, J. Da crise regional às novas dinâmicas de especialização da produção de cacau no Sul da Bahia. Caminhos de Geografia, v. 22, n. 79, p. 77–96, 1 fev. 2021b. <https://doi.org/10.14393/rcg227953461>.

XAVIER, L. B.; NASCIMENTO JR, F. D. C. do; CHIAPETTI, J. Verticalidades e horizontalidades na certificação da produção de cacau orgânico no Sul da Bahia. Revista Campo-Território, v. 15, n. 38, p. 361–390, 23 abr. 2021a. <https://doi.org/10.14393/rct153814>.