

Método Ensemble de Modelos Híbridos para Detecção de Barragem de Rejeitos em Imagem de Satélite

Diego A. B. Cardoso, Matheus R. Guimarães, Roberta B. Oliveira

* Departamento de Ciência da Computação, Universidade de Brasília,
DF (e-mails: diegodi05@hotmail.com, matheusrg007@gmail.com,
roberta.oliveira@unb.br)

Abstract: In Brazil, there is a high rate of irregular dams, including places that are not even cataloged. These dams have a high potential for socio-environmental damage and a possible rupture would be capable of destroying rivers and villages. This paper proposes an ensemble method composed of hybrid models, based on Convolutional Neural Network (CNN) and Support Vector Machine (SVM) for detecting tailings dams in satellite imagery. The database used was *BrazilDAM*, considering 1924 multispectral images, where some combinations were evaluated using 13 different bands. The proposed method achieved 96.1% accuracy, 96.8% F1-score, and 98.7% AUC. The application of the proposed method shows better results than the traditional methods proposed in the literature.

Resumo: No Brasil, há um índice alto de barragens irregulares, incluindo locais até mesmo não catalogados. Essas barragens têm alto potencial de dano socioambiental, considerando que um possível rompimento de suas estruturas seria capaz de destruir rios e vilarejos. O presente trabalho propõe um método *ensemble* composto por modelos híbridos, baseados em *Convolutional Neural Network (CNN)* e *Support Vector Machine (SVM)*, para a tarefa de detecção de barragens de rejeito a partir de imagens de satélite. A base de dados utilizada foi a *BrazilDAM* com a avaliação de algumas combinações das 13 diferentes bandas, considerando 1924 imagens multiespectrais. O método proposto obteve acurácia de 96,1%, *F1-score* 96,8% e *AUC* 98,7%. Os resultados apresentam que a aplicação do método proposto traz melhores resultados para a detecção de barragens de rejeitos em comparação aos métodos propostos na literatura.

Keywords: convolutional neural network; support vector machine; ensemble; hybrid model; tailings dam.

Palavras-chaves: rede neural convolucional; máquina de vetores de suporte; ensemble; modelo híbrido, barragem de rejeito.

1. INTRODUÇÃO

As barragens de rejeitos são grandes reservatórios que são compostos de água e resíduos sólidos provindos da atividade de mineração (Luz et al., 2010). Este tipo de estruturas pode representar grandes riscos, pois possuem uma capacidade de causar danos ambientais em caso de um eventual rompimento. Os trágicos acidentes ocorridos em 2015 e 2019 nas cidades mineiras de Mariana e Brumadinho, respectivamente, demonstram concretamente quão elevado é o potencial de danos causados por alguma falha nas barragens (Pereira et al., 2021). Outro problema, que torna ainda mais prejudicial, são as barragens em situação irregular pelos mais diversos motivos, tais como, ausência de informação da altura da fundação, sem apresentar coordenadas e até mesmo não possuem registros de nome. Segundo a reportagem realizada pela Agência Brasil¹, o Brasil possuía em 2018 cerca de 45% de suas barragens

catalogadas em situação irregular. A Agência Nacional de Mineração (ANM)² possui um controle a respeito de diversas características, tais como risco, potencial de dano e localização de barragens. Além das barragens reguladas e catalogadas pelo governo, existem também barragens ilegais onde não é feito nenhum tipo de levantamento sobre os impactos ambientais causados por tal instalação.

A tarefa de detecção de barragens tem recebido uma atenção especial, com destaque para os últimos anos, como demonstram os autores Balaniuk et al. (2020) que utilizam Redes Neurais Convolucionais (*Convolutional Neural Networks - CNNs*) para identificação e classificação de barragens em imagens de satélite. Além disso, devido ao tamanho das tragédias citadas anteriormente, há alguns estudos que consideram base de dados de barragens brasileiras, como é o caso dos autores Ferreira et al. (2020) que propõem uma base de dados com imagens de satélites das barragens catalogadas no Brasil.

¹ <https://agenciabrasil.ebc.com.br/geral/noticia/2018-04/especialistas-debatem-seguranc-e-fiscalizacao-de-barragens-no-pais>. Acesso em: 16/05/2022

² <https://www.gov.br/anm/pt-br>

Tendo em vista o problema de localizar possíveis barragens não catalogadas, há algumas maneiras para tentar solucionar tal problema, por exemplo, considerando técnicas para a tarefa de classificação de imagens. Modelos baseados em Aprendizado Profundo têm sido utilizados na literatura para diferentes problemas relacionados à classificação de imagens, como mostram os autores Ferreira et al. (2020) que utilizam CNN. Os modelos híbridos também estão sendo aplicados para classificação de imagens. Este tipo de modelo consiste na fusão de diferentes técnicas, como por exemplo a fusão da CNN com a Máquina de Vetor de Suporte (Support Vector Machine - SVM), para a classificação de imagens (Tao and Wei, 2022). Além disso, também estão presentes na literatura os chamados *ensembles* de classificadores que, por sua vez, são a combinação de modelos de classificação para obtenção de melhores resultados em comparação ao classificador individual. Os ensembles podem ser construídos de diversas maneiras, como por exemplo, considerando a combinação de vários modelos de CNN (Turkoglu et al., 2022).

O objetivo do trabalho (Cardoso et al., 2022) que deu origem a este artigo era desenvolver um método *ensemble* baseado em modelos híbridos, a fim de auxiliar na localização de barragens ilegais. Sendo estes modelos compostos por uma CNN para extração de características e uma SVM para a tarefa de classificação de barragens em imagens de satélites. Para isso, foi necessário realizar uma sequência de experimentos que visam melhorar o desempenho dos modelos de CNNs utilizados na literatura para o problema em questão.

Na Seção 2 são apresentados trabalhos da literatura que motivaram a proposta das técnicas aplicadas no presente artigo. Na Seção 3 é detalhado o método proposto para a classificação de barragens de rejeitos em imagens de satélite. Na Seção 4 são abordados os experimentos realizados e os resultados obtidos em cada um deles. Na Seção 5 é realizada uma discussão geral sobre os resultados apresentados. Por fim, na Seção 6 são abordadas a conclusão do método para detecção de barragens e as propostas para trabalhos futuros.

2. TRABALHOS RELACIONADOS

Há alguns trabalhos na literatura que utilizam CNN para realizar a detecção de barragens de rejeitos a partir de imagens de satélite de alta resolução. No trabalho apresentado por Ferreira et al. (2020) é introduzida a base de dados *BrazilDAM* a qual possui imagens dos satélites *Sentinel-2* e *Landsat-8*. Além disso, os autores realizam um *Benchmark* entre as principais CNNs utilizadas na literatura para a tarefa de classificação de objetos em imagens. Nos experimentos realizados, as redes testadas foram *VGG16*, *AlexNet*, *DenseNet121*, *InceptionV3*, *SqueezeNet* e *ResNet50*, sendo que a *DenseNet121* obteve o melhor resultado com uma acurácia média de 94,11% para a tarefa de classificação binária da base de dados sendo barragem ou não-barragem. Os resultados foram calculados a partir da média da acurácia sobre as cinco validações cruzadas realizadas no conjunto de teste. Além disso, os autores indicaram o intervalo com 95% de confiança obtido. Para esta tarefa foi utilizado apenas o conjunto de dados com as imagens obtidas do satélite *Sentinel-2* no ano de 2019.

Este conjunto possui 1924 imagens, sendo 769 consideradas como barragem e 1156 consideradas não barragem.

Outra abordagem utilizada na literatura é o uso de uma Rede Neural Convolutacional Totalmente Conectada (*Fully Convolutional Network* - FCN) para a mesma tarefa de identificação e classificação mencionada anteriormente, como mostra o estudo de Balaniuk et al. (2020). Os autores propõem o uso de uma FCN para identificar e classificar regiões de barragens e minas a partir de coordenadas de áreas que contenham ou não barragens de rejeitos. Para gerar os dados, os autores fazem uso de uma técnica de raspagem de dados para buscar essas coordenadas através de uma Interface de Programação de Aplicações do *Google Earth*, e portanto, obter imagens multiespectrais dos locais indicados pelas coordenadas. Formando uma base de dados que possui 2860 imagens, sendo 1397 barragens ou minas e 1463 que não são. Além disso, os resultados alcançados pelo autores foram promissores, sendo a acurácia 97,44%, *F1-score* 97,49%, revocação 97,47% e Área Abaixo da Curva ROC³ (*Area Under the ROC Curve* - AUC) 97,46%.

Abrangendo mais um pouco a área e indo além do tema de detecção de barragens, mas apresentando uma técnica que inspirou o presente trabalho, os autores Tao and Wei (2022) propõem um modelo híbrido composto por uma CNN e uma SVM com objetivo de fazer reconhecimento a partir de imagens de ervas daninhas em campos no período de inverno. A arquitetura CNN utilizada foi a *VGG16* para realizar a extração de características das imagens. Já a SVM, com o *kernel* Função de Base Radial (*Radial Basis Function* - RBF), é utilizada para realizar a classificação das imagens. Nesse trabalho, para explicar a adoção do modelo *VGG*, ainda é proposta uma comparação entre os desempenhos dos modelos considerados estado-da-arte para a mesma tarefa. Os autores obtiveram como resultado uma acurácia de 92,1% na base de teste (240 imagens) utilizando o modelo híbrido proposto *CNN-SVM*.

O trabalho de Turkoglu et al. (2022) tem como objetivo detectar doenças e pragas em plantas utilizando a técnica de *ensemble*. Para isso, os autores utilizam cinco modelos de CNN, que são *AlexNet*, *GoogleLeNet*, *Densenet201*, *Resnet50* e *Resnet101*, sendo cada um deles utilizado para extração de características, e uma SVM para realizar a predição quanto à praga. Para definir a classificação final foi utilizada a votação majoritária entre os resultados obtidos pelos cinco modelos para o mesmo conjunto de dados de entrada. Como resultados foram obtidos uma acurácia de 97,56% e *F1-score* 97,07%.

O presente trabalho combina diversas técnicas discutidas na literatura para detecção de imagens, sendo que neste trabalho são utilizadas para o problema de detecção de barragens em imagem de satélite. Assim, o objetivo deste trabalho é propor um método capaz de obter resultados melhores do que os já apresentados na literatura para o problema em questão, e portanto, auxiliar na localização de barragens ilegais.

³ Curva ROC: Curva Característica de Operação do Receptor (Receiver Operating Characteristic Curve)

3. MÉTODO PROPOSTO

O método proposto no presente trabalho consiste em um *ensemble* de modelos híbridos baseados em CNN e SVM para a tarefa de classificação binária de barragens em imagens de satélite. A Figura 1 apresenta as etapas, técnicas e processos do método proposto. Nesta seção são elucidados em detalhes todas as etapas do método presente na Figura 1: (1) base de dados utilizada; (2) pré-processamento das imagens; (3) aplicação de *transfer learning* e *fine tuning* nas CNNs, e utilização da SVM como classificador, formando o modelo híbrido; (4) *ensemble* dos modelos da etapa anterior; e (5) etapa de votação majoritária para classificação final, considerando as classes Barragem e Não é Barragem.

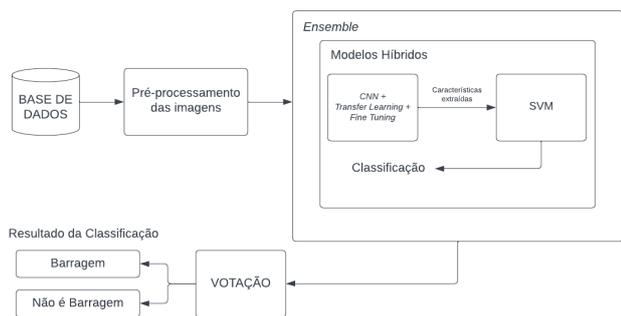


Figura 1. Visão geral do método proposto.

3.1 Base de Dados

A base de dados que foi utilizada no presente trabalho foi a *BrazilDAM* (Ferreira et al., 2020). Essa base possui imagens multiespectrais (384×384 pixels)⁴. Essas imagens foram obtidas por meio dos satélites *Sentinel-2*⁵ e *Landsat-8*⁶ e ambos possuem imagens da mesma região dos anos de 2016-2019 de barragens catalogadas do Brasil.

Para os experimentos serão considerados somente as imagens do ano de 2019 do *Sentinel-2*, pois foram avaliadas antes na literatura (Ferreira et al., 2020). Além disso, o uso de imagens da mesma região de diferentes anos poderia causar o enviesamento do modelo para a tarefa que está sendo trabalhada, conforme testes preliminares no presente trabalho indicaram. A base utilizada possui no total 1924 imagens com 13 bandas espectrais, na qual 769 imagens são de barragens e 1155 não são. Vale ressaltar que as imagens obtidas pelo satélite *Landsat-8* não foram utilizadas pois possui apenas oito bandas e, além disso, tais imagens seriam as mesmas obtidas pelo satélite *Sentinel-2*. A Figura 2 mostra exemplos de imagens em *RGB* da base de dados, a partir da combinação de bandas, nas quais as imagens da Figura 2(a) e (c) são da classe Barragem e da Figura 2(b) e (d) da classe Não é Barragem.

As imagens da base de dados *BrazilDAM* estão no formato de arquivo TIF e, portanto, possui além das imagens multiespectrais algumas informações técnicas como minério

⁴ Imagens que contêm diferentes comprimentos de ondas eletromagnéticas

⁵ <https://sentinel.esa.int/web/sentinel/missions/sentinel-2>. Acesso em: 16/05/2022

⁶ <https://www.usgs.gov/landsat-missions/landsat-8>. Acesso em: 16/05/2022

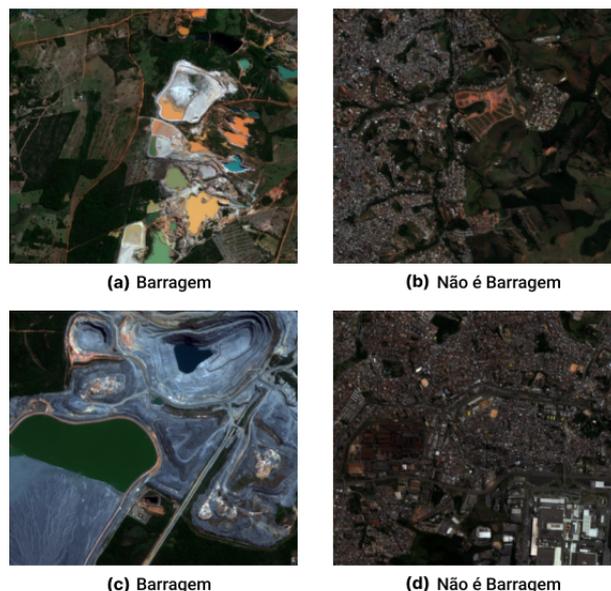


Figura 2. Exemplos de imagens de satélite (Ferreira et al., 2020).

principal, altura, volume, método construtivo, categoria de risco e dano potencial associado à barragem. Embora existam todas essas informações, para o presente trabalho somente serão utilizadas as imagens, pois o objetivo é apenas detectar as barragens.

3.2 Pré-processamento das Imagens

Trabalhar diretamente com arquivos TIF exige bastante poder computacional por conta das diversas informações que estão contidas neles como visto na Subseção 3.1. Além disso, como será feita a utilização das técnicas de *transfer learning* e *fine tuning*, trabalhar com imagens multiespectrais com 13 bandas não é o ideal por se diferenciar muito da base na qual os modelos serão pré-treinados (Yuan et al., 2021).

Portanto, foi considerada uma etapa de pré-processamento das imagens, com objetivo de trabalhar com as imagens em um formato mais convencional, como o PNG, no qual não são perdidas tantas informações em sua compressão (Wiggins et al., 2001) em comparação com alguns outros métodos. Para isso, foi utilizada somente três bandas para que a aplicação das técnicas citadas anteriormente possam acontecer da melhor maneira e assim esperar por melhores resultados. Além disso, é possível realizar uma análise de diferentes combinações dentre as 13 bandas.

Como mencionado na Subseção 3.1, o *Sentinel-2* possui 13 bandas espectrais que podem ser agrupadas da seguinte maneira⁷:

- **Bandas de 4×10 metros:** bandas *RGB* (*Red* - B4, *Green* - B3 e *Blue* - B2) e banda NIR (*Near Infrared* - B8);
- **Bandas de 6×20 metros:** bandas VNIR (*Visible Near Infrared* - B5, B6, B7, B8A) e bandas SWIR (*Short-Wave Infrared* - B11 e B12);

⁷ <https://sentinel.esa.int/web/sentinel/user-guides/sentinel-2-msi/resolutions/radiometric>. Acesso em: 16/05/2022

- **Bandas de 3 × 60 metros:** banda *Ultra blue (Coastal and Aerosol - B1)* e banda SWIR (B10).

A utilização de combinações de bandas, permite extrair informações específicas de uma imagem. Por exemplo, existem combinações de bandas que destacam características geológicas, agrícolas ou de vegetação em uma imagem⁸.

Para o presente trabalho foram geradas cinco versões da base de dados *BrazilDAM* utilizando o *software GDAL*⁹. Cada versão representa as combinações das diferentes bandas com base nos padrões presentes na aplicação *Sentinel Playground*¹⁰. Na Figura 3 podem ser vistos os resultados da combinação de diferentes bandas na mesma imagem da base de dados. Além disso, essas combinações foram nomeadas para as novas imagens pré-processadas que serão utilizadas na próxima etapa do presente trabalho, sendo: (a) *RGB*; (b) *SWIR*; (c) *Infrared*; (d) *Agriculture*; e (e) *Bathymetric*. Por fim, as imagens foram redimensionadas através do método de interpolação bilinear para o tamanho de 224 × 224 *pixels*.

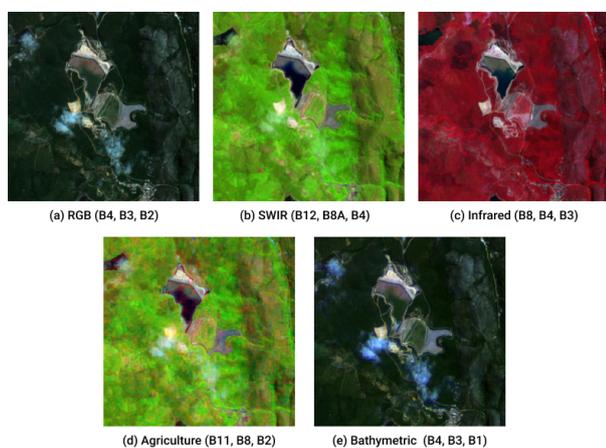


Figura 3. Demonstração de combinações de bandas.

3.3 Aplicação de Transfer Learning e Fine Tuning nas CNNs

As técnicas de *transfer learning* (Pan and Yang, 2010) e *fine tuning* (Peng and Wang, 2020) consistem em realizar um treinamento do modelo em uma base de dados diferente da base do problema com objetivo de que o modelo aprenda características genéricas como contornos e formas geométricas. Estas técnicas podem ser capazes de aumentar drasticamente os resultados de um modelo, principalmente quando a base de dados a ser utilizada não é tão grande. Portanto, para o presente trabalho foram utilizadas as técnicas citadas anteriormente em cada um dos três modelos de CNNs para a extração de características, as quais farão parte dos experimentos, *VGG16* (Liu and Deng, 2015), *DenseNet121* (Huang et al., 2016) e *InceptionV3* (Szegedy et al., 2015). Esses modelos foram considerados, pois apresentaram os melhores resultados no trabalho de Ferreira et al. (2020). Vale ressaltar que os modelos estão sendo utilizados com os parâmetros padrões

⁸ <https://gisgeography.com/sentinel-2-bands-combinations>. Acesso em: 16/05/2022

⁹ <https://gdal.org>

¹⁰ <https://www.sentinel-hub.com/explore/sentinelplayground/>

definidos na biblioteca *keras* presente no *TensorFlow*¹¹, sendo a função de ativação *Sigmoid* utilizada para classificação.

Para a realização do *transfer learning* foram utilizados os pesos a partir da aplicação dos modelos na base de dados *ImageNet* (Russakovsky et al., 2014), que é uma base muito utilizada na literatura, e conta com mais de 1 milhão de imagens distribuídas em 1000 classes. A base *ImageNet* foi considerada por conta da sua variedade de dados o que pode contribuir para o aprendizado de características genéricas do modelo. Os pesos da base *ImageNet* estão disponibilizados na biblioteca *keras*.

Já no *fine tuning*, primeiro foi executado cada modelo mencionado anteriormente por dez épocas considerando os subconjuntos de treinamento e de validação, que são abordados na Seção 4, com a aprendizagem de todas as suas camadas congeladas. Após isso foram executadas mais vinte épocas deixando somente o equivalente a 20% das camadas descongeladas, para aprenderem as especificidades da base de dados *BrazilDAM*, totalizando 30 épocas. Vale observar que estas camadas são as últimas de cada modelo de CNN.

Para a primeira execução do *fine tuning* foi utilizado como otimizador o *Adam* (Kingma and Ba, 2015) com um *learning rate* de 0,0001. Já na segunda execução do *fine tuning* o otimizador utilizado foi *RMSprop* (Tieleman et al., 2012), com *learning rate* de 0,00001. Para as duas execuções, a função *loss* utilizada foi *Binary Cross Entropy* (Ruby and Yendapalli, 2020), pois o problema considerado no presente trabalho é de classificação binária, e portanto, o parâmetro de métrica utilizada foi a acurácia. Além disso, foi utilizada a técnica de *data augmentation* por meio do método *ImageDataGenerator* do *keras* para o treinamento dos modelos de CNNs. Essa técnica não foi utilizada com a finalidade de aumentar a quantidade de imagens da base, mas sim para trazer diversidade durante o treinamento, utilizando operações aleatórias de *zoom*, *distorção no sentido anti-horário* e *giro horizontal* da imagem.

3.4 Modelos Híbridos

No presente trabalho foi desenvolvido um modelo híbrido para a classificação, que é composto por CNN e SVM. A Figura 4 mostra a estrutura do modelo híbrido, que consiste em uma fusão de modelos visando obter melhores resultados em comparação com a utilização de apenas um modelo (Jena et al., 2021). Essa abordagem tem mostrado resultados promissores na literatura (Tao and Wei, 2022).

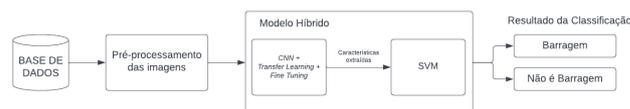


Figura 4. Estrutura do modelo híbrido CNN-SVM.

Diante disso, a CNN foi considerada apenas para extrair as características da imagem. Foram removidas das CNNs as suas camadas de classificação que passaram pelo processo de *transfer learning* e *fine tuning* como citado na

¹¹ <https://www.tensorflow.org/>. Acesso em: 16/05/2022

Subseção 3.3. Além disso, diferentes modelos foram avaliados, como será abordado na Seção 4, com o objetivo de usar o que obtiver melhor resultado para compor o modelo híbrido. Dessa forma, as imagens pré-processadas na etapa anterior foram transformadas em vetores de características, considerando o espaço de cor *RGB*. Após esse processo, os vetores resultantes são utilizados como dados de entrada para a SVM que realiza a classificação, neste caso como Barragem ou Não é Barragem. Foram utilizados os valores padrões da biblioteca *Scikit-learn*¹² para os parâmetros da SVM, com exceção da função *kernel* que será explorada na Subseção 4.2.

3.5 Ensemble de Classificadores

Ensembles são a combinação de múltiplos modelos de classificação que utilizam o resultado de cada classificador em benefício da solução do problema (Zhang and Ma, 2012). Pode-se pensar que cada classificador individual que compõem o *ensemble* é como um especialista em algo e o resultado final é gerado a partir da combinação dos resultados de cada especialista.

Uma parte importante das arquiteturas *ensemble* é a combinação dos modelos pois é onde os resultados serão convergidos em resultado final. Para o presente trabalho foi utilizado a estratégia de voto majoritário (Zhang and Ma, 2012), que consiste em cada modelo dar o resultado da sua classificação e o resultado da maioria é o resultado final. Existem diversas técnicas para a realização de *ensemble*. No presente trabalho foram avaliadas duas variações, sendo elas uma homogênea e uma heterogênea (Zhang and Ma, 2012).

No caso do *ensemble* homogêneo, o conjunto de classificadores utilizados para realizar a tarefa de classificação pode ser idêntico. O classificador poderá receber um conjunto de dados diferentes, conseqüentemente, gerando diferentes modelos de acordo com quantidade de subconjuntos considerados. Desta forma, o classificador pode aprender uma característica específica do seu conjunto e contribuir de forma diferente no resultado. A Figura 5 apresenta o *ensemble* homogêneo considerado no presente trabalho o qual é constituído por modelos híbridos conforme visto na Subseção 3.4. O modelo possui diferentes versões da base de dados as quais estão presentes na Figura 3 e explicadas na Subseção 3.2. Estas bases alimentam instâncias diferentes do mesmo classificador. Os subconjuntos de teste de todas as as versões da base de dados são agregados para a criação da base de testes, que será explicada na Seção 4, e será utilizada para a realização da classificação no módulo de votação. Este módulo combina todos os cinco modelos e, realiza a votação majoritária para as classes barragem ou não barragem.

Já no caso do *ensemble* heterogêneo, o conjunto de classificadores pode conter diferentes tipos de arquitetura. Nesse caso, a base de dados que é utilizada por cada um deles é a mesma, sendo assim cada arquitetura irá contribuir com a sua especialidade. A Figura 6 apresenta um exemplo de estrutura *ensemble* heterogêneo, que que é aplicado em uma mesma base de dados e que será definida na Seção 4, sendo uma das versões apresentadas na Subseção 3.2. Esta

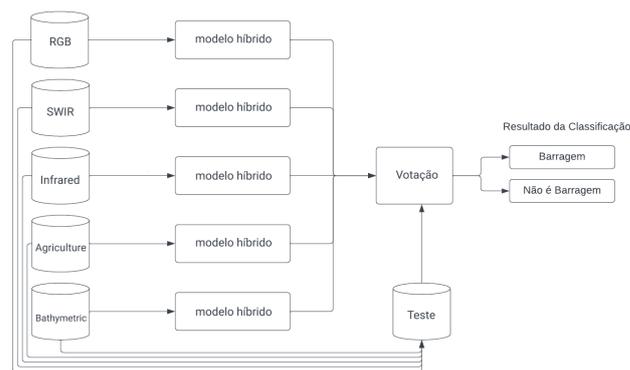


Figura 5. Estrutura do modelo *ensemble* homogêneo.

base alimenta as três configurações propostas de modelos híbridos diferentes e, ao fim, as três configurações são convergidas em uma etapa que realiza uma votação majoritária resultando na classificação para Barragem ou Não é Barragem. Todas as configurações para o *ensemble* são apresentadas e avaliadas na Seção 4.

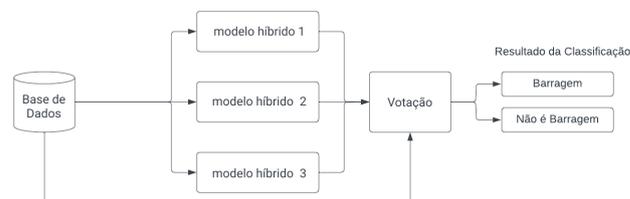


Figura 6. Estrutura do modelo *ensemble* heterogêneo.

4. RESULTADOS EXPERIMENTAIS

Para a realização dos experimentos relacionados ao método proposto, a base de dados que foi vista na Seção 3.1 foi dividida de forma aleatória e estratificada, ou seja, com proporção similar entre as classes por partição em subconjuntos de treino, validação e teste. Para treino foram utilizados 70% dos dados, o equivalente a 1346 imagens, e para validação e teste 15% cada, o equivalente a 289 imagens. Os conjuntos de treino e validação são utilizados para o desenvolvimento dos modelos e o conjunto de teste para a avaliação final. Para realizar a avaliação dos experimentos presentes nesta seção, serão utilizadas as seguintes métricas (Hossin and Sulaiman, 2015): Acurácia, Precisão, Revocação, *F1-score* e AUC.

4.1 Experimento I: CNNs utilizando Transfer Learning e Fine Tuning

O Experimento I tem como objetivo comparar os resultados da *VGG16*, *DenseNet121* e *InceptionV3*, considerando a técnica de *Transfer Learning* e com a utilização do *Fine Tuning*. A partir do experimento foi possível entender que com o *transfer learning* o modelo é capaz de alcançar resultados promissores para a tarefa de classificação de imagens de barragem. Além disso, foi possível analisar se a aplicação do *Fine Tuning* consegue melhores resultados em comparação com o que foi alcançado sem o uso da técnica. Para as etapas de treino, validação e teste foi utilizada a versão da base de dados no espaço de cor *RGB*, conforme visto na Subseção 3.2, por ser mais comum na literatura.

¹² <https://scikit-learn.org/>. Acesso em: 16/05/2022

Considerando a utilização apenas da técnica de *Transfer Learning*, é possível observar a partir da Tabela 1 que o modelo *InceptionV3* alcançou uma Acurácia, Revocação e *F1-score* melhores em comparação com os demais modelos. Os resultados de Acurácia e Revocação da *InceptionV3* indicam que o modelo está realizando melhores classificações no cenário geral. Além disso, o resultado para *F1-score* indica que os resultados são promissores para classificação positiva (Barragem). Por outro lado, a *Densenet121* obteve a melhor AUC, o que significa que consegue separar melhor as classes Barragem e Não é Barragem em comparação com as demais. Já a Precisão indica que o resultado das imagens classificadas como barragem contém menos Falsos Positivos do que o resultado obtido pela Inception. Para esta comparação, a *VGG16* teve os piores resultados em relação aos demais modelos.

Tabela 1. Comparação entre CNNs.

Modelos	Acurácia	Precisão	Revocação	F1-Score	AUC
<i>Transfer Learning</i>					
VGG16	0,792	0,846	0,797	0,821	0,888
DenseNet121	0,820	0,895	0,791	0,840	0,923
InceptionV3	0,844	0,829	0,930	0,877	0,906
<i>Transfer Learning + Fine Tuning</i>					
VGG16	0,906	0,905	0,942	0,923	0,952
DenseNet121	0,923	0,946	0,924	0,935	0,960
InceptionV3	0,889	0,893	0,924	0,909	0,953

A Tabela 1 demonstra que de fato a aplicação da técnica *Fine Tuning* alcançou melhoras significantes para quase todas as métricas avaliadas nos modelos. Apenas a Revocação da *InceptionV3* foi ligeiramente pior em relação à utilização apenas do *Transfer Learning*. A *VGG16*, que obteve os piores resultados anteriormente, foi capaz de aumentá-los de tal forma que os seus resultados foram superiores aos resultados da aplicação apenas do *Transfer Learning*. Inclusive, foi a que obteve a melhor Revocação entre os modelos, ou seja, de todas as imagens que são Barragens previu corretamente a classificação como sendo Barragem. Em contrapartida, a *Densenet121* foi a que mais aproveitou os benefícios advindos da técnica aplicada, conseguindo resultados promissores para Acurácia, Precisão, *F1-score* e AUC.

4.2 Experimento II: Modelo Híbrido

Para este experimento foi utilizado o modelo de CNN que obteve o melhor resultado no experimento da Subseção 4.1, que foi aplicando *Transfer Learning* e *Fine Tuning*. Dito isto foi desenvolvido um modelo híbrido utilizando a CNN para extração de características e a SVM para realizar a classificação. Portanto, um dos objetivos é descobrir se o modelo híbrido terá bons resultados em relação aos modelos de CNN vistos no experimento anterior. Outro objetivo é analisar qual função *kernel* melhor se encaixa para o modelo.

A partir da Tabela 2, é evidenciado que o modelo híbrido de fato foi capaz de melhorar em relação aos resultados dos modelos de CNNs, tendo em vista que todas as métricas receberam um aumento significativo. Além disso, é possível observar que a função *kernel* linear também obteve os melhores indicativos.

Tabela 2. Resultados da variação da função *kernel* no modelo híbrido.

Kernel SVM	Acurácia	Precisão	Revocação	F1-Score	AUC
Linear	0,941	0,948	0,953	0,951	0,970
RBF	0,930	0,927	0,959	0,943	0,980
Polinomial	0,923	0,912	0,965	0,938	0,976
Sigmoidal	0,937	0,937	0,959	0,948	0,976

4.3 Experimento III: Combinação de bandas espectrais

Baseado nos resultados do experimento da Subseção 4.2 o melhor modelo híbrido constatado foi a combinação *DenseNet121* e SVM com a função *kernel* linear. Até então o espaço de cor que estava sendo utilizado em todos os experimentos era o *RGB*, portanto, o objetivo deste experimento é verificar se a mudança de bandas espectrais, vista na Subseção 3.2, influencia no resultado do modelo híbrido.

A Tabela 3 traz os resultados de todas as combinações de bandas espectrais aplicadas no modelo híbrido *DenseNet121-SVM*. Os resultados para *RGB*, *SWIR* e *Agriculture* foram bem próximos. Sendo que os três modelos, inclusive, chegaram no mesmo resultado para *F1-score* e Acurácia. O *RGB* obteve a maior precisão que foi de 0,948, o *Agriculture* obteve a maior revocação que foi de 0,971 e o *SWIR* obteve a maior AUC que foi de 0,981.

Tabela 3. Resultados dos diferentes espaços de cores no modelo híbrido.

Espaço de Cor	Acurácia	Precisão	Revocação	F1-Score	AUC
RGB	0,941	0,948	0,953	0,951	0,970
SWIR	0,941	0,943	0,959	0,951	0,981
InfraRed	0,927	0,928	0,945	0,936	0,976
Agriculture	0,941	0,933	0,971	0,951	0,966
Bathymetric	0,889	0,880	0,933	0,905	0,952

Considerando a diferença entre as outras métricas, o *SWIR* com a AUC obteve a maior distância de valores em comparação com os valores de Precisão e Revocação. A partir da métrica AUC é possível observar que a versão da base de dados considerando a combinação de bandas *SWIR* lidou melhor com a separação das classes Barragem ou Não é Barragem em comparação com as outras combinações de espaços de cores.

4.4 Experimento IV: Ensembles Homogêneos e Heterogêneos

O experimento da Subseção 4.3 mostrou que a mudança de espaço de cor realmente influencia no resultado do modelo híbrido. Portanto, este experimento tem como objetivo testar se a construção de um ensemble a partir da variação de subconjuntos de entrada considerando diferentes versões da base de dados é capaz de melhorar ainda mais os resultados. Para isso são utilizadas cinco instâncias do modelo híbrido *DenseNet121-SVM* com a função *kernel* linear, definida na Subseção 4.2, sendo cada uma treinada com uma variação da base de dados, ou seja, diferentes combinações de bandas conforme abordado na Subseção 3.5. Para também avaliar se o modelo híbrido funcionará bem em conjunto com o *ensemble homogêneo*, foi feita uma comparação utilizando *ensembles* dos modelos híbridos *Densenet121-SVM* e das CNNs *Densenet121*.

Os resultados para o subconjunto de teste são apresentados na Tabela 4.

Tabela 4. Comparação de *ensemble* homogêneo de CNNs e de modelos híbridos.

Ensemble Homogêneo	Acurácia	Precisão	Revocação	F1-Score	AUC
CNNs	0,892	0,909	0,908	0,908	0,927
Modelos Híbridos	0,913	0,890	0,971	0,929	0,975

É evidente que o *ensemble* de modelos híbridos teve resultados superiores em comparação com o *ensemble* de CNNs, como verifica-se na Tabela 4. Todavia, os modelos *ensembles* homogêneos não conseguiram aumentar os resultados em comparação com os experimentos da Subseção 4.3. Isto demonstra que o conhecimento adquirido para trabalhar com um espaço de cor diferente do qual foi treinado por cada instância não funcionou bem. Além disso, a técnica de *ensemble* homogêneo exige mais recursos para sua execução já que precisa de múltiplas bases de dados para seu treinamento.

O *ensemble* homogêneo não conseguiu melhorar os resultados anteriores. Portanto, outra possibilidade é considerar um modelo *ensemble* heterogêneo, para verificar se a combinação de diferentes modelos de classificação são capazes de melhorar os resultados. Para isso, a base de dados utilizada é a com o espaço de cores SWIR, pois ela obteve os melhores resultados no Experimento III, e são considerados o *VGG16*, o *DenseNet121* e o *InceptionV3*. Desta mesma forma, é comparado neste experimento os *ensembles* de CNNs e de modelos híbridos. Os resultados para o subconjunto de teste são apresentados na Tabela 5.

Tabela 5. Comparação de *ensemble* heterogêneo de CNNs e de modelos híbridos.

Ensemble Heterogêneo	Acurácia	Precisão	Revocação	F1-Score	AUC
CNNs	0,951	0,954	0,965	0,959	0,984
Modelos Híbridos	0,961	0,950	0,988	0,968	0,987

A Tabela 5 traz resultados interessantes da utilização do *ensemble* heterogêneo baseado em modelos híbridos. Mais uma vez, como esperado, o modelo híbrido apresentou melhores resultados em relação ao modelo com CNNs. Outro ponto importante é que o modelo *ensemble* heterogêneo obteve os melhores resultados em todas as métricas em comparação com todos os outros experimentos anteriores. Sendo assim, é possível concluir que o *ensemble* heterogêneo com modelos híbridos para a tarefa de classificação de barragens possui um resultado promissor em comparação com os outros modelos utilizados na literatura.

4.5 Ambiente Experimental

O desenvolvimento, treinamento e execução de todos os experimentos foram realizados no ambiente *Google Colaboratory Pro* em um computador com processador *Intel(R) Xeon(R) CPU @ 2.30GHz*, placa de vídeo *Tesla P100-PCIE-16GB*, 25GiB de memória *RAM* e *HD 125GiB*. Além disso, foi utilizada a linguagem de programação *Python*, as bibliotecas *keras* presente no *TensorFlow* e o *Scikit-learn*. Além disso, foi utilizada a biblioteca *python opencv2*¹³ para manipulação das imagens. Todos os experimentos levaram em torno de 10 horas e 47 minutos para

¹³<https://opencv.org>

serem finalizados. Vale ressaltar que esse tempo contempla somente a leitura da base e execução do treinamento, validação e teste sendo obtido a partir de uma única execução.

5. DISCUSSÃO DOS RESULTADOS

Por meio dos resultados obtidos pelos experimentos realizados na seção anterior, é possível identificar uma série de fatores que são relevantes para o problema de classificação de barragens em imagens de satélites. Utilizando apenas o *transfer learning* para avaliação do modelo foi possível obter bons resultados já que o procedimento realizado foi apenas definir as CNNs e pré-carregar os pesos da base *ImageNet*. Ao final, a *InceptionV3* se mostrou com melhores resultados em comparação com a *Densenet121* e a *InceptionV3*. Já com a utilização do *Fine Tuning* é notório um aumento significativo nos resultados, o que mostra o quanto a técnica funcionou bem para o problema aqui estudado. Porém, como a *InceptionV3* não conseguiu melhorar os resultados nas mesmas proporções que os demais modelos, pode-se verificar que a técnica em questão funciona melhor em determinados modelos.

Foi possível identificar que a utilização de modelos híbridos é capaz de melhorar ainda mais os resultados obtidos em comparação com apenas o uso de CNNs. Além disso, ficou evidenciado que a função *kernel* linear se adaptou melhor ao modelo híbrido junto a uma *Densenet121*. As combinações de espaço de cores se mostram capazes de influenciar o modelo, embora não tão significativamente quanto o esperado. Apesar de que para os olhos humanos a combinação de outras bandas de cores pareça tão diferente, aparentemente, para o modelo abordado não há tanta distinção no processo de aprendizado. Embora a técnica *ensemble* visar aumentar os resultados alcançados por modelo de classificação individual, o modelo *ensemble* homogêneo desenvolvido no presente artigo não foi capaz de aumentar os resultados em comparação com modelo híbrido.

O *ensemble* heterogêneo de modelos híbridos se mostrou o melhor em comparação com os outros modelos testados no presente trabalho, conseguindo resultados promissores de acurácia 0,961, precisão 0,950, revocação 0,988, *F1-score* 0,968 e AUC 0,987. Isto demonstra que este modelo pode atuar bem para problemas de classificação de barragens a partir de imagens de satélites. A desvantagem é que a técnica utiliza vários modelos, fazendo com que a demanda de tempo para execução seja maior, já que todos precisam ser treinados. Além disso, o método proposto no presente trabalho obteve melhores resultados em comparação com o estudo apresentado por Ferreira et al. (2020). Apesar de que as metodologias e abordagens de avaliação foram diferentes do estudo comparado.

Como limitações do trabalho, devido à falta de recursos computacionais pessoais não foi possível utilizar a técnica de *k-fold Cross-Validation*, que foi a técnica mais utilizada na literatura para o problema em questão e que também é interessante para avaliar base de dados pequenas (Refaeilzadeh et al., 2009). Além disso, não foram feitos experimentos para ajustes de hiper-parâmetros, uma vez que isto poderia incrementar os resultados obtidos. Por fim, os modelos utilizados foram treinados e testados apenas na base de dados *BrazilDAM* (Ferreira et al., 2020), não sendo

possível ter uma visão mais ampla do comportamento do modelo em outras bases de dados de imagens de barragens.

6. CONCLUSÃO E TRABALHOS FUTUROS

O presente artigo apresentou uma associação de diversas técnicas como *transfer learning*, *fine tuning*, modelo híbrido CNN-SVM e *ensemble* de modelos de classificação, que ao serem combinadas culminaram em um método que mostrou resultados bastante positivos. Este método *ensemble* proposto, por sua vez, obteve os melhores resultados em comparação aos modelos tradicionais e híbridos. Com acurácia 0,961, precisão 0,950, revocação 0,988, F1-score 0,968 e AUC 0,987. Sendo assim, *ensemble* de modelos híbridos para a tarefa de classificação de barragens em imagens de satélite se mostrou bastante promissor.

Mesmo com os bons resultados do presente trabalho, há experimentos que poderiam avaliar a capacidade de melhorias para elevar ainda mais o potencial do método proposto. Como diferentes maneiras de combinação dos classificadores para o modelo *ensemble* ou o uso de outras técnicas tradicionais de *ensemble*: Bagging, Boosting ou Stacking (Zhang and Ma, 2012). Outro ponto possível de teste seria o uso de outras combinações de espaços de cores, dado que o *Sentinel-2*¹⁴ possui 13 bandas espectrais. Para testar todas estas combinações possíveis resultariam em 286 diferentes versões da base no qual, para a tarefa aqui avaliada, alguma destas combinações poderia influenciar o resultado. Além disso, realizar um estudo mais profundo em relação aos parâmetros e hiper-parâmetros utilizados, tanto na SVM como nas CNNs.

REFERÊNCIAS

- Balaniuk, R., Isupova, O., and Reece, S. (2020). Mining and tailings dam detection in satellite imagery using deep learning. *Sensors*, 20(23).
- Cardoso, D.A.B., Guimarães, M.R., and Oliveira, R.B. (2022). Ensemble de modelos híbridos baseados em redes neurais convolucionais e máquinas de vetores de suporte para detecção de barragens em imagens de satélite. Trabalho de conclusão de curso (Bacharelado em Ciência da Computação). Universidade de Brasília (UnB).
- Ferreira, E., Brito, M., Alvim, M.S., et al. (2020). Brazil-dam: A benchmark dataset for tailings dam detection. In *Latin American GRSS ISPRS Remote Sensing Conference 2020, Santiago, Chile*. IEEE.
- Hossin, M. and Sulaiman, M. (2015). A review on evaluation metrics for data classification evaluations. *International Journal of Data Mining Knowledge Management Process*, 5, 01–11.
- Huang, G., Liu, Z., and Weinberger, K.Q. (2016). Densely connected convolutional networks. *CoRR*, abs/1608.06993.
- Jena, B., Saxena, S., Nayak, G.K., Saba, L., Sharma, N., and Suri, J.S. (2021). Artificial intelligence-based hybrid deep learning models for image classification: The first narrative review. *Computers in Biology and Medicine*, 137, 104803.
- Kingma, D.P. and Ba, J. (2015). Adam: A method for stochastic optimization. *Proceedings of the 3rd International Conference on Learning Representations (ICLR 2015)*.
- Liu, S. and Deng, W. (2015). Very deep convolutional neural network based image classification using small training sample size. In *2015 3rd IAPR Asian Conference on Pattern Recognition (ACPR)*, 730–734.
- Luz, A.B.d., Sampaio, J.A., and França, S.C.A. (2010). *Tratamento de minérios*. 5.ed. Rio de Janeiro: CETEM/MCT.
- Pan, S.J. and Yang, Q. (2010). A survey on transfer learning. *IEEE Transactions on Knowledge and Data Engineering*, 22(10), 1345–1359.
- Peng, P. and Wang, J. (2020). How to fine-tune deep neural networks in few-shot learning? *CoRR*, abs/2012.00204.
- Pereira, D.M., Guimarães, H.O.R., Freitas, S.M.C.d., and Mângia, A.A.M. (2021). Brumadinho: Muito mais que um desastre tecnológico. *Revista da Universidade Federal de Minas Gerais*, 27(2), 332–355.
- Refaeilzadeh, P., Tang, L., and Liu, H. (2009). *Cross-Validation*, 532–538. Springer US, Boston, MA.
- Ruby, U. and Yendapalli, V. (2020). Binary cross entropy with deep learning technique for image classification. *International Journal of Advanced Trends in Computer Science and Engineering*, 9.
- Russakovsky, O., Deng, J., Su, H., et al. (2014). Imagenet large scale visual recognition challenge. *International Journal of Computer Vision*, 115.
- Szegedy, C., Vanhoucke, V., Ioffe, S., et al. (2015). Rethinking the inception architecture for computer vision. *CoRR*, abs/1512.00567.
- Tao, T. and Wei, X. (2022). A hybrid CNN-SVM classifier for weed recognition in winter rape field. *Plant Methods*, 18.
- Tieleman, T., Hinton, G., et al. (2012). Lecture 6.5-RMSprop: Divide the gradient by a running average of its recent magnitude. *COURSERA: Neural networks for machine learning*, 4(2), 26–31.
- Turkoglu, M., Yanikoğlu, B., and Hanbay, D. (2022). Plantdiseasenet: convolutional neural network ensemble for plant disease and pest detection. *Signal, Image and Video Processing*, 16(2), 301–309.
- Wiggins, R.H., Davidson, H.C., Harnsberger, H.R., et al. (2001). Image file formats: Past, present, and future. *RadioGraphics*, 21(3), 789–798.
- Yuan, K., Zhuang, X., Schaefer, G., et al. (2021). Deep-learning-based multispectral satellite image segmentation for water body detection. *IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing*, 14, 7422–7434.
- Zhang, C. and Ma, Y. (2012). *Ensemble Machine Learning: Methods and Applications*. Springer Publishing Company, Incorporated.

¹⁴ <https://sentinel.esa.int/web/sentinel/user-guides/sentinel-2-msi/resolutions/radiometric>