

Explorando o Potencial da IA Generativa para Melhorar a Classificação de Estilos Arquitetônicos com Data Augmentation

Caio Assunção Albuquerque, Yuri De Almeida Malheiros Barbosa, Thaís Gaudencio do Rêgo

Centro de Informática – Universidade Federal da Paraíba (UFPB)

João Pessoa – PB – Brazil

1.caioalbuquerque@gmail.com, yuri@ci.ufpb.br, gaudenciothais@gmail.com

Resumo: O interesse crescente na inteligência artificial generativa destaca-se como uma oportunidade promissora para ampliar os limites da criação e manipulação de imagens. Especificamente, os modelos de difusão oferecem novas perspectivas para aumentar tanto o volume quanto a diversidade das imagens, que são fundamentais para treinar modelos precisos de classificação. Neste estudo, buscamos explorar o potencial do modelo *Stable Diffusion* para enriquecer os conjuntos de dados de treinamento e, conseqüentemente, aprimorar modelos de classificação de imagens, com foco na identificação de estilos arquitetônicos de monumentos históricos. Essa abordagem permitiu a geração de novas imagens a partir de prompts de texto e uma imagem de entrada, proporcionando um aumento significativo no volume do conjunto de dados de treinamento. Além disso, para treinar os modelos de classificação de imagens, foi empregado uma rede neural convolucional (CNN), ResNet50, utilizando o conjunto de dados expandido, composto pelas imagens sintéticas e pelas imagens originais. Foram realizados 5 experimentos. No experimento 1, foi explorado o potencial das imagens sintéticas na melhoria da capacidade de generalização do classificador de estilo Barroco, mantendo uma proporção de 50% para cada tipo de imagem, neste experimento, não foi usado *data augmentation*, e foi observado pequenas variações na acurácia, sugerindo que as imagens sintéticas podem ser utilizadas para o *data augmentation* sem perder as características essenciais de cada classe. Nos experimentos 2 e 3, o *data augmentation* foi aplicado apenas a uma única classe de desempenho abaixo da média, triplicando o volume de imagens do conjunto de treino, enquanto no experimento 4, foi realizado *data augmentation* para as duas classes de pior desempenho. Para esses três experimentos houve melhorias na taxa de acerto do classificador onde o conjunto de treino foi aumentado. Já no experimento 5, foi feito *data augmentation* para todas as classes, aumentando em até 4 vezes o volume de imagens geradas para cada classe, entretanto não houve mudanças significativas na acurácia das classes, exceto para classe neoclássica, neste houve uma melhoria significativa na acurácia, de 60,34% para 82,76%, embora tenha sido notada uma tendência em classificar erroneamente imagens ecléticas como neoclássicas.

Palavras-chave: Stable Diffusion, Aumento de dados, IA generativa, Rede Neural Convolucional.

Abstract: The growing interest in generative artificial intelligence stands out as a promising opportunity to expand the boundaries of image creation and manipulation. Specifically, diffusion models offer new perspectives to increase both the volume and diversity of images, which are crucial for training accurate classification models. In this study, we aimed to explore the potential of the Stable Diffusion model to enrich training datasets and consequently enhance image classification models, focusing on identifying architectural styles of historical monuments. This approach enabled the generation of new images from text prompts and an input image, leading to a significant increase in the volume of the training dataset. Additionally, to train the image classification models, a convolutional neural network (CNN), ResNet50, was employed using the expanded dataset composed of synthetic and original images. Five experiments were conducted. In Experiment 1, we explored the potential of synthetic images to improve the generalization ability of the Baroque style classifier while maintaining a 50% proportion for each image type. Data augmentation was not used in this experiment, and slight variations in accuracy were observed, suggesting that synthetic images can be used for data augmentation without losing the essential characteristics of each class. In Experiments 2 and 3, data augmentation was applied only to a single below-average performing class, tripling the volume of images in the training set, while in Experiment 4, data augmentation was performed for the two worst-performing classes. For these three experiments, there were improvements in the classifier's accuracy where the training set was augmented. In Experiment 5, data augmentation was performed for all classes, increasing the volume of generated images for each class up to four times; however, there were no significant changes in the accuracy of the classes except for the Neoclassical class. In this class, there was a significant improvement in accuracy, from 60.34% to 82.76%, although a tendency to misclassify eclectic images as neoclassical was noted.

Keywords: Stable Diffusion, Data Augmentation, Generative AI, Convolutional Neural Network.

Catálogo na publicação
Seção de Catalogação e Classificação

A345e Albuquerque, Caio Assunção.

Explorando o potencial da IA generativa para
melhorar a classificação de estilos arquitetônicos com
data augmentation / Caio Assunção Albuquerque. - João
Pessoa, 2024.

41 f. : il.

Orientação: Thaís Gaudêncio do Rêgo.

Coorientação: Yuri De Almeida Malheiros Barbosa.
TCC (Graduação) - UFPB/CI.

1. Inteligência artificial generativa. 2. Aumento de
dados. I. Gaudêncio do Rêgo, Thaís. II. De Almeida
Malheiros Barbosa, Yuri. III. Título.

UFPB/CI

CDU 004.8

1 Introdução

Uma das principais subdivisões da inteligência artificial (IA) é o aprendizado de máquina, composto por diversos métodos construídos com base em modelos matemáticos derivados da teoria estatística. Tais métodos capacitam os computadores a executar tarefas automaticamente, ao identificar padrões presentes nos conjuntos de dados existentes ou nas experiências passadas [1]. Um subcampo especializado em aprendizagem de máquina é o chamado aprendizado profundo, que busca reformular as representações derivadas dos dados por meio de uma abordagem única. É ressaltado nessa abordagem o processo de aprender por meio de camadas sucessivas. Nesse contexto, essas representações em camadas costumam ser feitas utilizando Redes Neurais Artificiais [2].

Apesar dos avanços contínuos, as redes neurais profundas ainda dependem de grandes quantidades de dados de treinamento para evitar o superajustamento do modelo (*overfitting*). No entanto, dados rotulados para aplicações do mundo real, como dados usados para classificar tipos arquitetônicos de imagens de monumentos históricos, são limitados e com políticas rigorosas de direitos autorais, tornando a construção da sua base de dados mais difícil [3]. Além disso, o problema de desequilíbrio de dados, por exemplo onde alguns estilos arquitetônicos possuem significativamente menos amostras que outras, também pode afetar o desempenho dos modelos de classificação. Portanto, as técnicas existentes de aumento de dados (AD) representam uma das maneiras eficazes de melhorar a quantidade e a diversidade dos dados de treinamento.

Os exemplos mais bem-sucedidos de AD são encontrados na aplicação a conjuntos de dados constituídos por imagens. Nestas situações, os padrões originais podem ser modificados através da aplicação de transformações geométricas, como rotação, ampliação, translação, variação de contraste, entre outras, sem que isso afete a classificação das amostras [4]. Entretanto, muitas vezes o aumento de dados não conseguem melhorar a precisão da classificação dos modelos de aprendizagem profunda (*deep learning*), e por esse motivo os modelos generativos como Redes Adversariais Generativas (do inglês, *Generative Adversarial Network* - GAN) e modelos de difusão podem ser usados para tentar superar essas limitações. Esses modelos oferecem novas abordagens para a geração de imagens sintéticas e são eficazes em diversas aplicações [5].

Neste contexto, esse estudo propõe utilizar dos modelos de difusão, os quais permitem a criação de novas imagens, a partir de uma imagem base e *prompts* de texto, para o aumento de dados dos cinco principais estilos arquitetônicos do Brasil: Barroco, Neoclássico, Neogótico, Eclético e Modernista. Foi utilizado o conceito de difusão como base, que é capaz de transformar gradualmente amostras de ruído em imagens detalhadas e de alta qualidade, proporcionando uma maneira eficaz de expandir e enriquecer conjuntos de dados de imagens relacionadas à arquitetura brasileira, trazendo benefícios para a pesquisa e fornecendo resultados para discutir a viabilidade do uso de aumento de dados através dos modelos de difusão.

2 Modelos de Difusão e Trabalhos Relacionados

2.1 Modelos de Difusão

Os modelos de difusão foram popularizados nos últimos anos para geração de texto para imagem, como também de imagem para imagem [6]. Muitos modelos como Dall-E da OpenAI, Imagen do Google e Stable Diffusion da Stability AI são capazes de gerar imagens realistas de alta resolução. Existem diferentes tipos de arquiteturas de modelos de difusão, como modelos de difusão com eliminação de ruído, modelos de difusão latente, modelos de difusão baseados em pontuação e assim por diante [6]. Todos eles se baseiam no princípio de recuperar uma imagem de sua versão ruidosa obtida pela adição progressiva de ruído, como gaussiano [6]. A Figura 1 oferece uma ilustração conceitual do processo de difusão, proporcionando uma compreensão visual dos mecanismos subjacentes.

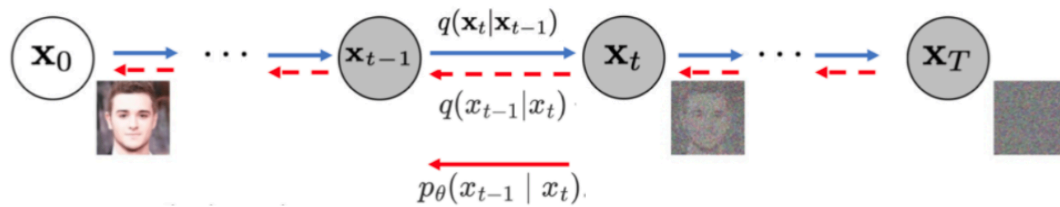


Fig. 1. O modelo probabilístico de difusão com eliminação de ruído envolve um processo direto conhecido como difusão e um processo reverso denominado difusão reversa ou eliminação de ruído. O processo de difusão (denotado como $q(x_t|x_{t-1})$) adiciona ruído, como o gaussiano, aos dados de forma incremental. Já no processo de difusão reversa (denotado como $q(x_{t-1}|x_t)$), esse ruído é estimado e revertido para recuperar a distribuição original dos dados. Na etapa de eliminação de ruído $p_\theta(x_{t-1}|x_t)$ as amostras são restauradas para sua forma original, completando o ciclo do modelo, indo do ruído aos dados novamente [7].

É importante ressaltar que, assim como outros modelos, os modelos de difusão também apresentam limitações significativas que precisam ser consideradas. Uma das limitações mais problemáticas diz respeito à eficiência computacional durante a inferência [8]. Isso ocorre devido ao alto número de etapas de avaliação necessárias para gerar uma única amostra. Essa ineficiência de tempo pode ser uma barreira significativa, especialmente em cenários onde a geração de amostras precisa ser rápida. Superar essa limitação sem comprometer a qualidade das amostras geradas representa uma área crucial para futuras pesquisas [8].

2.2 Modelo de Difusão Estável (Stable Diffusion)

O Modelo de difusão estável emprega aprendizagem profunda para a geração de imagens com base em texto, superando modelos convencionais baseados em GAN, ao produzir imagens mais fotorrealistas, criativas e diversificadas, atendendo com mais precisão aos critérios definidos pelo usuário durante o processo de geração [9]. Ao comprimir imagens em um espaço latente e gradualmente introduzir ruído, o *Stable Diffusion* é treinado para regenerar imagens, tornando-se uma solução eficaz na geração de imagens a partir de descrições textuais. Além disso, sua versatilidade permite não apenas a geração de imagens a partir de texto, mas também a realização de

traduções de imagem para imagem guiadas por texto [10]. Em resumo, o *Stable Diffusion* pode funcionar em ambos os domínios, convertendo texto em imagem e também imagem em imagem, desde que haja um contexto textual para orientar o processo de geração [11].

2.3 Trabalhos Relacionados

Nesta seção, vamos examinar pesquisas importantes nesta área, focando em novidades e obstáculos enfrentados em diversas abordagens metodológicas.

Recentemente, Sukkar e colaboradores (2024) abordaram o uso de sistemas generativos de IA, em particular o Midjourney, para criar imagens criativas de edifícios e locais do patrimônio arquitetônico islâmico a partir de texto. Eles destacam o potencial dessas representações geradas por IA para arquitetos, especialistas e usuários comuns, mas também examinam suas limitações ao gerar imagens que se afastam muito das estruturas originais representadas. O estudo utiliza técnicas de engenharia de *prompts* e comparações visuais diretas para avaliar a precisão das imagens geradas pelo *Midjourney*, em relação aos exemplos selecionados da tradição arquitetônica islâmica. Os autores concluem que, embora o *Midjourney* possa criar imagens de alta qualidade inspiradas na tradição islâmica, ainda não consegue reproduzir com precisão a aparência real de algumas estruturas originais [12].

Zhou e colaboradores (2023) propuseram um aumento de dados de eletroencefalografia (EEG) baseada em modelos de difusão (Diff-EEG) para melhorar o diagnóstico de Alzheimer. Nos experimentos, foi usado o EEG-Net e os resultados mostram que o método Diff-EEG obteve a maior precisão, alcançando 89,3% para um conjunto de dados de escala 2 e 93,3% para um conjunto de dados de escala 3, superando outros métodos, como Noisy, VAE e DCGAN. Os autores concluíram que o *framework* de aumento de dados, baseado em modelos de difusão proposto, apresenta um desempenho promissor na tarefa de diagnóstico de Alzheimer, especialmente quando combinado com a Spatial VQ-VAE, para extrair informações de relacionamento entre diferentes canais de EEG [13].

Ngasa e colaboradores (2024) desenvolveram um modelo generativo, denominado Rede Adversarial Generativa de Wasserstein com Penalidade de Gradiente (WGAN-GP) baseada em difusão, para gerar imagens de alta qualidade de glóbulos brancos (WBC). O modelo demonstrou sua eficácia em um conjunto de dados composto por 4.503 imagens em 19 classes de leucócitos do Hospital Bucheon da Universidade Soonchunhyang, na Coreia, mostrando uma melhoria significativa na geração de imagens de alta qualidade para classes raras de leucócitos e no tratamento do desequilíbrio de dados. A metodologia inclui a combinação das redes neurais convolucionais (do inglês, *Convolutional Neural Network* - CNNs) pré-treinados com Máquinas de Vetor de Suporte (do inglês, *Support Vector Machines* - SVM) para classificação, na qual a estratégia de aumento levou o modelo *ResNet50-SVM* a atingir uma precisão média de 95% na classificação das 19 classes WBC [5].

Moreno e Colaboradores (2023) apresentaram uma nova abordagem para detecção de ervas daninhas usando classificadores de CNN, especificamente Yolov8l e RetinaNet, aumentados com dados de *Stable Diffusion*, ou seja, imagens artificiais de ervas daninhas. O estudo teve como alvo ervas daninhas específicas encontradas em culturas de tomate, utilizando um número limitado de

imagens reais (30 amostras), para produzir imagens artificiais de treinamento para as CNNs. O desempenho dos modelos de detecção foi aumentado em 6% - 9% aumentando o conjunto de dados com imagens artificiais, atingindo uma média de precisão média (mAP) máxima de 93% e até 99%, quando apenas uma espécie foi detectada. Este estudo demonstra o potencial para utilização de *Stable Diffusion* no desenvolvimento de sistemas baseados em visão artificial para práticas agrícolas aprimoradas [14].

Borji (2022) realizou uma comparação quantitativa de três sistemas populares, *Stable Diffusion*, *Midjourney* e *DALL-E 2* para averiguar quão bem eles funcionam na geração de rostos fotorrealistas. O autor avaliou a qualidade das faces geradas em cenas desordenadas contendo múltiplos objetos e utilizou modelos geradores de texto em imagem para sintetizar cenas. Em seguida, foi usado um detector facial para detectar rostos nessas imagens. Por fim, foi utilizada a Distância de Iniciação de Fréchet (FID) para avaliar a qualidade das faces geradas, em relação a um conjunto de faces reais. O autor concluiu que o *Stable Diffusion* gera faces mais realistas do que os outros sistemas. Na maioria dos casos, porém, é possível para um ser humano dizer se um rosto é real ou gerado, indicando que ainda há uma grande lacuna a preencher [15].

Stöckl (2022) realizou análises para avaliar a capacidade do modelo *Stable Diffusion* de gerar imagens sintéticas corretas para uma grande variedade de conceitos, utilizando a taxonomia Wordnet e as definições de conceitos contidas nela. O autor observou que o *Stable Diffusion* gerou pelo menos uma imagem correta para 10 tentativas, de uma ampla gama de conceitos (73%), indicando que há informações no sistema para grande parte dos conceitos. No entanto, foi observado que as imagens geradas para um mesmo conceito podem ser bastante diferentes entre si, o que complica a tarefa do procedimento de classificação [16].

É importante destacar que essa pesquisa toma como ponto de partida um estudo anterior conduzido por Dias (2023), que explorou o reconhecimento de estilos arquitetônicos do Brasil por meio de um aplicativo móvel utilizando uma CNN, a *EfficientNet*, uma vez que a mesma é conhecida pela sua capacidade de alcançar resultados competitivos com menos recursos computacionais, em comparação com outras arquiteturas convolucionais mais complexas. Isso é crucial em dispositivos móveis, onde os recursos de processamento e memória são limitados. Sendo assim, ao optar pela *EfficientNet*, os desenvolvedores podem garantir uma experiência mais fluida aos usuários do aplicativo, mantendo a precisão e a eficácia do reconhecimento de padrões [27]. Esse trabalho mostrou-se relevante na esfera educacional, classificando e identificando os cinco principais estilos arquitetônicos, incluindo Barroco, Neoclássico, Neogótico, Eclético e Modernista [17].

Com o objetivo de ampliar esta linha de pesquisa, o presente estudo explora técnicas de aumento de dados para enriquecer a diversidade do conjunto de dados e aprimorar o reconhecimento dos estilos arquitetônicos. Utilizando como base o mesmo conjunto de dados que Dias (2023), optamos pela *ResNet* (*Residual Network*) como CNN, visando aumentar a precisão dos classificadores. É relevante destacar que a CNN neste estudo é utilizada para o treinamento do modelo, e ambos os estudos adotam o *framework TensorFlow*. Enquanto Dias (2023) realizou um aumento de dados convencional, com edições básicas nas imagens existentes, como zoom, rotação e flip, nossa abordagem se concentra na geração de imagens sintéticas adicionais utilizando o modelo *Stable Diffusion*.

Em seu estudo, Dias (2023) conduziu testes utilizando técnicas como: espelhamento, rotação, deslocamento, cisalhamento, zoom, variação de brilho e outras transformações para aumentar a diversidade dos dados de treinamento. Observou-se que o modelo demonstrou um bom desempenho na categoria Neogótico, com uma sensibilidade de 85%, indicando sua habilidade em identificar características distintivas desse estilo. Por outro lado, o estilo Eclético apresentou uma precisão de apenas 64%, destacando a complexidade inerente a esse estilo e a dificuldade do modelo em classificá-lo com precisão. Embora os estilos Barroco e Neoclássico tenham alcançado pontuações-F1 de 0,84 e 0,73, respectivamente, há espaço para melhorias, especialmente na sensibilidade do estilo Neoclássico [17].

Ao comparar os trabalhos relacionados com o estudo em questão, podem ser identificadas algumas semelhanças e diferenças significativas. Em primeiro lugar, é importante destacar que vários desses estudos utilizam modelos de difusão, como o *Stable Diffusion* e *Midjourney*, para a geração de imagens sintéticas em diferentes contextos. Por exemplo, Moreno et al. (2023), Stöckl (2023), Sukkar et al. (2024) exploram geração de imagens sintéticas de ervas daninhas, imagens gerais e monumentos da arquitetura islâmica, respectivamente, enquanto o estudo em questão se concentra na geração de imagens de estilos arquitetônicos presentes no Brasil.

Zhou et al. (2024) e Ngasa et al. (2024) exploram o aumento de dados em diferentes contextos médicos: o primeiro concentra-se no diagnóstico da doença de Alzheimer por meio de EEG, enquanto o segundo trata da classificação de imagens de glóbulos brancos. Zhou et al. (2024) empregam modelos de difusão guiada e *autoencoders* (VQ-VAE) para expandir os conjuntos de dados de EEG. Enquanto isso, Ngasa et al. (2024) utilizam WGAN-GP e DDPM para gerar imagens sintéticas de glóbulos brancos e melhorar a classificação, com o SVM sendo empregado para essa tarefa. Apesar de todos os estudos mencionados empregarem técnicas de aumento de dados, suas aplicações específicas e áreas de estudo diferem, demonstrando a versatilidade dessas técnicas de aumento de dados em áreas de pesquisa distintas.

Outra semelhança importante é a utilização de métricas de avaliação do desempenho do modelo, como acurácia, precisão, *recall* e pontuação-F1, em diversos trabalhos, incluindo o estudo em questão. Essas métricas permitem uma avaliação objetiva da capacidade do modelo de classificar corretamente. Todas essas comparações entre os trabalhos estão detalhadas na Tabela 1, apresentada a seguir, facilitando a análise e a compreensão das semelhanças e diferenças entre eles.

Tabela 1. Trabalhos Relacionados.

Autor	Objetivo	Algoritmo	Dataset	Métricas
SUKKAR et al., 2024 [12]	Definir as limitações atuais do sistema de IA do Midjourney na geração de imagens artificiais semelhantes ao patrimônio arquitetônico islâmico.	TextControlGAN	Imagens retiradas aleatoriamente da internet de monumentos islâmicos.	<i>Visual Attention Software</i> (VAS) e Realização de um inquérito com uma população selecionada para avaliar a semelhança entre as imagens geradas por IA e fotografias reais.
ZHOU et al., 2024 [13]	Desenvolver uma abordagem para aumentar os dados de eletroencefalografia (EEG) usando modelos de difusão, a fim de melhorar o diagnóstico da doença de Alzheimer	VQ-VAE e modelos de difusão guiada	O conjunto de dados de EEG utilizado neste trabalho foi obtido do artigo de Miltiadous et al. (2023)	Acurácia, precisão e <i>recall</i>
NGASA et al., 2024 [5]	Desenvolver uma ferramenta para um modelo generativo capaz de gerar imagens de glóbulos brancos (GB) em uma ampla gama de classes	WGAN-GP, DDPM e SVM	4.503 imagens de 19 classes de glóbulos brancos, provenientes do Hospital da Universidade de Soonchunhyang Bucheon, Coreia.	Acurácia, pontuação-F1, precisão e <i>recall</i>
MORENO et al., 2023 [14]	Desenvolver um novo método que utiliza o algoritmo de Difusão Estável para gerar imagens sintéticas de alta qualidade de ervas daninhas para a Gestão Específica de Ervas Daninhas (SSWM)	<i>Stable Diffusion</i>	Conjunto de dados misto com imagens reais de ervas daninhas	Média das precisões médias (mAP) e a taxa de processamento de imagens por segundo (FPS)
BORJI et al., 2022 [15]	Realizar uma comparação quantitativa de três sistemas populares de síntese de imagens - <i>Stable Diffusion</i> , <i>Midjourney</i> e <i>DALL-E 2</i> - em sua capacidade de gerar rostos fotorealistas	<i>Stable Diffusion</i> , <i>Midjourney</i> e <i>DALL-E 2</i>	Para as faces reais utilizaram o conjunto de treinamento COCO e adicionaram faces do conjunto de dados <i>Labeled Faces in the Wild</i> (LFW) e para faces geradas utilizaram legendas do conjunto de dados COCO como prompts para sintetizar imagens	FID (<i>Fréchet Inception Distance</i>)

STÖCKL, 2023 [16]	Gerar um conjunto de imagens sintéticas usando o modelo <i>Stable Diffusion</i> , empregando a taxonomia Wordnet, e explorar a utilidade dessas imagens geradas sinteticamente como dados de treinamento em aplicações de aprendizado de máquina	<i>Stable Diffusion</i> , implementado com a biblioteca <i>Diffusers</i> da <i>Huggingface</i>	A base de dados foi coletada a partir de imagens do Discord server do <i>Stable Diffusion</i> , criadas a partir de um prompt de texto para cada.	Taxa média de reconhecimento, calculada em função da profundidade das classes na taxonomia do Wordnet
DIAS, 2023 [17]	Desenvolver um aplicativo móvel que utilize uma CNN para simplificar o processo de reconhecimento e compreensão dos cinco principais estilos arquitetônicos do Brasil: Barroco, Neoclássico, Neogótico, Eclético e Modernista	CNN <i>EfficientNetV2B0</i>	Seleção de imagens representativas dos cinco principais estilos arquitetônicos presentes no Brasil: Barroco, Neoclássico, Neogótico, Eclético e Modernista. Essas imagens foram compiladas a partir de diferentes fontes, incluindo literatura acadêmica, bancos de dados arquitetônicos e opiniões de especialistas.	Matriz de confusão, acurácia, pontuação-F1

3 Metodologia de Pesquisa

Nesta seção, são descritos os passos realizados no trabalho. Inicia-se pela base de dados utilizada, onde é especificada sua composição e origem. Em seguida, detalha-se o processo de geração de novas imagens, explicando as tecnologias e metodologias empregadas para criar imagens sintéticas.

Na segunda etapa, são abordadas as CNNs e apresenta-se a arquitetura adotada para o treinamento dos modelos de classificação. Detalham-se aspectos relacionados à configuração das camadas e ao processo de treinamento desses modelos.

Na etapa subsequente, são destacadas as métricas selecionadas para avaliar o desempenho dos modelos. Em seguida, é explicado o desenho experimental, detalhando os experimentos planejados para investigar os efeitos das diferentes abordagens e técnicas empregadas ao longo do estudo, com o objetivo de compreender os resultados dos modelos de cada uma delas.

3.1 Base de dados

Os dados utilizados nesta pesquisa são da mesma base de dados do estudo intitulado: “Explorando estilos arquitetônicos: usando redes neurais aprimoradas para facilitar o aprendizado em dispositivos móveis” [17]. Essa base contém um total de 1449 imagens, divididas em cinco categorias de estilos arquitetônicos: Barroco, Eclético, Neoclássico, Neogótico e Modernista, como contabilizado na Tabela 2. São imagens retiradas da internet e de publicações que não possuem direitos autorais. A Figura 2 apresenta exemplos representativos de monumentos para cada estilo arquitetônico.



Fig. 2. Estilos arquitetônicos usados [17].

Tabela 2. Número de imagens por estilo arquitetônico.

Estilo Arquitetônico	Conjunto de Treino	Conjunto de Teste
Barroco	260	65
Neoclássico	226	58
Neogótico	200	50
Eclético	232	58
Modernista	240	60

Entre esses estilos, na arquitetura, são retomados em contraposição ao Barroco, o período dos “neos” – neoclássico, neogótico, etc [18]. Entretanto, o Barroco e o Neogótico compartilham certas características ornamentais e dramáticas. Além disso, tanto o Barroco quanto o Neoclássico enfatizam a ornamentação detalhada e a grandiosidade arquitetônica, valorizando a simetria e a proporção. Enquanto o Neoclássico retorna às formas da antiguidade clássica, o Barroco adota uma abordagem mais dramática e emotiva. Por outro lado, o Neogótico, com suas altas torres, arcos pontiagudos e uso extensivo de pedra, contrasta com o Modernismo, que prioriza linhas limpas, funcionalidade e uma abordagem minimalista desprovida de ornamentação. Essas semelhanças e diferenças podem ser exploradas para melhorar o desempenho dos modelos de classificação, com maior foco em refinamentos nos critérios de diferenciação entre os estilos mais distintos, como o modernista e o neoclássico, visando aumentar a assertividade na identificação das características específicas de cada um.

Devido à quantidade limitada de monumentos existentes na base de dados, optamos por empregar o conceito de aumento de dados utilizando IA generativa. Especificamente, utilizamos a técnica de *Stable Diffusion*, com a expectativa de que essa abordagem seja capaz de potencializar os classificadores dos modelos com menor taxa de assertividade. Acreditamos que essa estratégia permitirá a geração de um conjunto mais diversificado e robusto de imagens, auxiliando no treinamento dos modelos e, conseqüentemente, melhorando seu desempenho na classificação de estilos arquitetônicos.

3.2 Geração de novas imagens

Inicialmente, foram realizadas as importações necessárias, incluindo bibliotecas como *Torch*, *PIL* e *Diffusers*. Em seguida, o *Stable Diffusion (versão 1.5)* foi utilizado como parte de um *pipeline* de geração de imagens, com o auxílio da biblioteca *Diffusers*. Além disso, foram ativadas algumas configurações adicionais no *pipeline*, como o *attention slicing* e o *memory efficient* do *XFormers*.

Dos parâmetros empregados para geração, vale ressaltar dois, **seed** e **strenght**. O **seed** é um número usado para inicializar o gerador de números aleatórios, assegurando a reprodutibilidade dos resultados entre diferentes execuções. Ele garante que a mesma sequência de números seja gerada, facilitando a comparação e análise dos resultados. Na Figura 3 é possível observar as variações das imagens artificiais geradas. Por outro lado, o parâmetro **strength**, variando de 0 a 1, controla a intensidade da difusão, afetando diretamente a clareza e nitidez das imagens produzidas. Ajustar adequadamente esses parâmetros é crucial para obter resultados desejáveis e controlar a qualidade

das imagens geradas. Após uma extensa série de testes, foi constatado que o processo de geração de imagens no *Stable Diffusion* é fortemente influenciado pelo parâmetro **strength**. Ele determina a quantidade de ruído adicionada à imagem inicial, sendo que quanto maior o seu valor, maior será o nível de ruído introduzido. Quando a intensidade atinge seu máximo, configurado como 1, o ruído adicionado também é máximo, resultando em uma imagem final que não mais se baseia na imagem inicial, apenas no *prompt* de texto. Nesse cenário, o processo de redução de ruído é executado ao longo do número especificado de iterações de inferência, ignorando completamente a imagem de entrada (original). Na Figura 4, apresenta-se a imagem original, enquanto na Figura 5 são exibidas as transformações decorrentes do aumento do parâmetro **strength**. À medida que esse parâmetro é aumentado, observa-se que a imagem tende a se distanciar progressivamente da sua representação real, adquirindo características mais abstratas e menos definidas. Esse fenômeno é resultado da ampliação da variação introduzida pela IA generativa, que busca explorar novas possibilidades dentro do espaço latente, ao ponto que pode enfatizar ou perder características arquitetônicas da imagem de entrada.



Fig. 3. Variação do parâmetro seed (100 - 500)



Fig. 4. Fotografia de monumento arquitetônico barroco de entrada



Fig. 5. Variação do parâmetro **strength** (0,5 - 0,9)

Ao criar os prompts, inicialmente optou-se por algumas opções mais generalistas, como *'Photography of baroque architecture buildings'*, *'Photograph of a historic monument in the neo-gothic architectural style'* e *'Architectural detail of a neo-gothic church, showcasing its ornate decorations'*, todas adaptadas para o estilo arquitetônico retratado na imagem de entrada. Contudo, para uma abordagem mais refinada, desenvolveu-se prompts adicionais que exploram características específicas e distintivas dos diferentes estilos arquitetônicos. Por exemplo, *'complex facade of a neoclassicist monument'*, destacando suas proporções harmoniosas e características clássicas ou *'Detailed image of the facade of a Baroque style church, with ornamental details, very bright colors and a square without cars'*, ressaltando características barrocas na imagem. Essas orientações mais específicas incentivam a realçar elementos distintivos e detalhes característicos de cada estilo.

As imagens sintéticas foram geradas e selecionadas manualmente, implicando que cada imagem foi escolhida criteriosamente por avaliação humana. Este procedimento foi repetido para cada estilo arquitetônico apresentado. Essa abordagem manual tornou o processo mais trabalhoso, exigindo um entendimento detalhado das características específicas de cada classe arquitetônica. Como resultado, o procedimento demandou tempo e atenção especial para garantir a qualidade das imagens sintéticas produzidas, assegurando que elas capturassem as nuances e características distintivas de cada estilo arquitetônico.

3.3 Arquiteturas CNN

As CNNs têm se destacado em uma ampla gama de aplicações relacionadas ao processamento de imagens e sua categorização [19]. Seu funcionamento envolve a aplicação de filtros convolucionais na imagem de entrada, capturando padrões locais e produzindo mapas de características, que são passados para as camadas subsequentes da rede [20]. As primeiras camadas aprendem a identificar características simples, como bordas e texturas, enquanto camadas mais profundas são capazes de detectar padrões mais complexos e abstratos, como objetos e formas. Um ponto importante a ser observado é que as CNNs são frequentemente combinadas com outras técnicas, como *pooling*, normalização e ativação não linear, o que aumenta ainda mais sua capacidade de aprendizado e generalização [21].

A *AlexNet*, uma CNN que constitui de cinco camadas convolucionais e duas camadas totalmente conectadas, foi uma das primeiras redes a se destacar na classificação de imagens. Após a *AlexNet*, várias outras CNNs influentes surgiram, incluindo *Network-In-Network* (NIN), *VGG-Net*, *GoogLeNet*, *ResNet* e *DenseNet* [22]. Dentre essas, as que se concentram em aumentar a profundidade do modelo para melhorar a estrutura da rede são a *VGG-Net*, *GoogLeNet* e *ResNet*.

ResNet é conhecida por sua capacidade de treinar redes profundas sem o problema do desaparecimento dos gradientes, que é um problema comum em redes muito profundas. Isso é alcançado através da introdução de conexões residuais, que permitem que a rede aprenda o mapeamento residual entre a entrada e a saída de uma camada, facilitando o fluxo de gradientes durante o treinamento [23]. Essas conexões residuais ajudam a mitigar a degradação do desempenho que geralmente ocorre com o aumento da profundidade da rede, permitindo a construção de modelos mais profundos. O artigo original sobre o *ResNet* propôs cinco tamanhos diferentes do modelo: 18, 34, 50, 101 e 152 camadas. Desde então, muitas outras variantes do *ResNet* foram desenvolvidas, como *ResNeXt* e *Wide Residual Networks* (WRN). Por exemplo, o *ResNet-34* usa uma arquitetura de rede simples inspirada no VGG-19, adicionando conexões de atalho e mais poderosas. Essa capacidade de treinar redes mais profundas levou a melhorias significativas no desempenho em uma variedade de tarefas de visão computacional e reconhecimento de padrões [23].

3.4 Criação do modelo

Para iniciar esse processo, foi adotada uma estratégia de aprendizagem por transferência, começando com o modelo *ResNet50*. Reconhecido por sua ampla aplicabilidade e robustez, o *ResNet50* é pré-treinado e consolidado, oferecendo vantagens significativas, como profundidade de rede, desempenho confiável e resolução eficaz de problemas de desaparecimento de gradientes. Desde 2015, tem sido amplamente utilizado em aprendizagem por transferência, com seu desempenho comprovado em diversas aplicações [25].

Em sequência, foi implantada uma série de camadas adicionais para refinamento e customização do processo de aprendizado. Inicialmente, uma camada de *Global Average Pooling* 2D foi utilizada para reduzir a dimensionalidade dos dados e extrair características globais relevantes das imagens processadas [26]. Em seguida, quatro camadas *Dense* foram adicionadas, cada uma composta por 1024 unidades neurais ativadas pela função de ativação *ReLU* (*Rectified Linear Unit*), visando introduzir complexidade e capacidade de aprendizado ao modelo [27]. Consequentemente, outra camada *Dense* foi aplicada com 512 unidades neurais ativadas pela função *softmax*, responsável por gerar as previsões finais do modelo para as cinco classes distintas de imagens. Essa arquitetura de camadas permite ao modelo capturar e aprender características discriminativas importantes das imagens de entrada, resultando em previsões precisas e mais confiáveis.

Durante o processo de treinamento de um modelo de CNN, é comum empregar técnicas como o ajuste fino (*fine-tuning*) de modelos pré-treinados para melhorar o desempenho em tarefas específicas [28]. Uma estratégia amplamente adotada é congelar as camadas iniciais do modelo para preservar as características aprendidas durante o treinamento prévio. Esse procedimento visa aproveitar os recursos de baixo nível aprendidos pelo modelo base, ao mesmo tempo em que adapta partes específicas do modelo para a nova tarefa.

Para garantir que as imagens de entrada sejam adequadamente pré-processadas, antes de serem alimentadas ao modelo durante o treinamento, foi utilizado o *TensorFlow* (v2.15.0) e sua classe **ImageDataGenerator**. Essa classe é configurada com uma função de pré-processamento específica, neste caso, *tf.keras.applications.resnet50.preprocess_input*, que realiza a normalização dos valores dos pixels das imagens para garantir que estejam na faixa esperada pela rede neural. Assim, asseguramos o tratamento adequado das imagens de entrada durante o treinamento.

Por fim, o modelo foi compilado utilizando a função de perda de entropia cruzada categórica e métricas de precisão. Essas diretrizes estabelecem como os pesos são atualizados e como a performance é avaliada durante o treinamento, utilizando a diferença entre as distribuições de probabilidade previstas e as distribuições reais das classes, além de medir a proporção de previsões corretas. O treinamento foi conduzido ao longo de um total de 80 épocas, com critério de parada antecipada, focado na monitoração da perda nos dados do conjunto de teste, para otimizar o desempenho sem superajustamento.

O primeiro modelo foi criado sem o uso de nenhuma técnica de aumento de dados (*data augmentation*). Posteriormente, foram realizados ajustes na base de dados para avaliar os impactos dessas modificações nos resultados dos classificadores. Esses ajustes podem sugerir melhorias em determinados cenários.

3.5 Métricas

As métricas de desempenho são úteis para comparar a qualidade das previsões entre sistemas. Algumas métricas comumente usadas para classificação binária são exatidão, precisão, *recall*, pontuação-F1 e uma **cálculo de proporção** no conjunto de treinamento [29]. Nesse estudo foram adotadas quatro métricas para análise dos resultados: a acurácia, a matriz de confusão e a pontuação-F1. A **acurácia (AC)** representa a porcentagem de previsões corretas em relação ao total de previsões. É uma métrica geral de desempenho do modelo. A **Matriz de confusão** é uma matriz de dimensão $n \times n$, onde n representa o número de classes a serem previstas. Seu propósito é calcular as métricas associadas à classificação, incluindo a contagem de falsos positivos (FP), falsos negativos (FN), verdadeiros positivos (*True Positives* - TP) e verdadeiros negativos (*True Negatives* - TN). A **pontuação-F1** combina precisão e *recall* em uma única medida, fornecendo uma avaliação mais abrangente do desempenho do modelo. Ele varia entre 0 e 1, onde 1 indica o melhor resultado possível, e é especialmente útil em conjuntos de dados desbalanceados, onde algumas classes têm mais dados de exemplos do que outras. No cálculo de proporção do conjunto de treinamento, dividimos o número de ocorrências do conjunto desejado pelo total de ocorrências em todos os conjuntos, visando identificar padrões ou tendências [29].

3.6. Desenho Experimental

Neste tópico, discutiremos os experimentos conduzidos no estudo, que se concentraram em explorar o potencial da geração de imagens sintéticas para aumentar os conjuntos de dados de treinamento e avaliar sua influência na eficiência do modelo.

3.6.1. Experimento 1: Uso de imagens sintéticas na classe estilo Barroco.

O experimento visa explorar o potencial das imagens sintéticas na melhoria da capacidade de generalização de um classificador, aqui no caso, do estilo Barroco. Substituímos metade das 260 imagens reais por imagens sintéticas, mantendo uma proporção de 50% para cada tipo. Ao introduzir essa variabilidade no conjunto de dados, o classificador foi desafiado a generalizar padrões de maneira mais robusta. Essa estratégia visa explorar o potencial das imagens sintéticas (Figura 6), proporcionando uma visão de como o classificador pode reagir à inclusão de imagens sintéticas, possibilitando uma avaliação precisa de seu desempenho e robustez em diferentes cenários. Isso é especialmente relevante em situações em que há escassez de imagens reais. A escolha de focar no estilo Barroco se deve à análise prévia dos resultados obtidos em trabalhos relacionados, onde as classes eclética e neoclássica apresentaram os piores desempenhos, enquanto o neogótico e os estilos modernistas já alcançaram resultados satisfatórios. Além disso, observou-se uma sugestão de melhoria na classificação do estilo Barroco, tornando a classe interessante para ser avaliada no experimento.



Fig. 6. Conjunto de imagens sintéticas de monumentos do estilo barroco

3.6.2. Experimento 2: Aumento de dados para coleção de treinamento do estilo arquitetônico Neoclássico

O objetivo do experimento é explorar o potencial do aumento de dados no treinamento do classificador de uma determinada classe, no caso, o estilo arquitetônico neoclássico. Para isso, foram utilizadas imagens sintéticas geradas por modelos de difusão, conforme ilustrado na Figura 7. Os detalhes sobre a quantidade de imagens sintéticas e reais estão presentes na Tabela 3. Essa estratégia busca enriquecer a diversidade e representatividade do conjunto de treinamento, permitindo que o modelo aprenda uma ampla variedade de características visuais associadas ao estilo neoclássico.

Analisando a Tabela 3, é evidente o desbalanceamento, sobretudo na classe Neoclássico, com uma proporção de aproximadamente 41.60%, enquanto as outras classes têm proporções em torno de 14%. Essa disparidade considerável entre a classe Neoclássico e as demais classes pode comprometer a capacidade do modelo de generalizar adequadamente para todas as classes, devido ao desbalanceamento das classes no conjunto de dados. Assim, o modelo pode estar superajustado para a classe majoritária, neoclássica.

Tabela 3. Número de imagens dos conjuntos e *Imbalance Ratio* por estilo arquitetônico.

Estilo Arquitetônico	Imagens reais	Imagens sintéticas	Conjunto de treino total	Conjunto de treino	Proporção (%)
Barroco	260	0	260	65	16,30
Neoclássico	226	448	664	58	41.60
Neo-Gótico	200	0	200	50	12,53
Eclético	232	0	232	58	14,54
Modernista	240	0	240	60	15,03



Fig. 7. Conjunto de imagens sintéticas do estilo neoclássico

3.6.3. Experimento 3: Aumento de dados para coleção de treinamento do estilo arquitetônico Eclético

O propósito deste experimento é investigar o potencial do aumento de dados no treinamento específico do classificador de estilo arquitetônico eclético. Para isso, foram empregadas imagens sintéticas geradas por modelos de difusão, como ilustrado na Figura 8. Os detalhes sobre a quantidade de imagens sintéticas e reais podem ser encontrados na Tabela 4. Essa abordagem visa

enriquecer a diversidade e representatividade do conjunto de treinamento, permitindo que o modelo aprenda uma ampla gama de características visuais associadas ao estilo eclético.

Analisando a Tabela 4, podemos observar que as classes com maior desbalanceamento são Neoclássico e Eclético, com proporções de aproximadamente 14.00% e 42.70%, respectivamente. Esse desbalanceamento pode levar a um superajustamento dessa classe majoritária e comprometer a capacidade do modelo de generalizar adequadamente para todas as classes.

Tabela 4. Número de imagens dos conjuntos e *Imbalance Ratio* por estilo arquitetônico.

Estilo Arquitetônico	Imagens reais	Imagens sintéticas	Conjunto de treino total	Conjunto de treino	Proporção (%)
Barroco	260	0	260	65	16,10
Neoclássico	226	0	226	58	14,00
Neo-Gótico	200	0	200	50	12,40
Eclético	232	458	690	58	42,70
Modernista	240	0	240	60	14,90



Fig.8. Conjunto de imagens sintéticas do estilo eclético

3.6.4. Experimento 4: O uso do aumento de dados para classes: neoclássico e eclético

O objetivo deste experimento é investigar o potencial do aumento de dados no treinamento de um modelo de classificação de estilos arquitetônicos, com foco nas classes neoclássico e eclético, que também interferem no estilo barroco devido a algumas características arquitetônicas semelhantes. Nos Experimentos 2 e 3, foi aplicado aumento de dados de forma exclusiva para essas classes,

respectivamente. Observou-se uma diminuição no desempenho do modelo em classificar corretamente essas classes. Diante disso, surgiu o interesse em realizar um aumento de dados utilizando imagens sintéticas exclusivamente para essas duas classes. Essa abordagem foi adotada para investigar o comportamento do novo modelo.

Observando as proporções na Tabela 5, para as classes Neoclássico e Eclético de aproximadamente 32.30% e 33.60%, respectivamente, fica evidente o desbalanceamento entre as classes. Essas duas classes apresentam proporções significativamente maiores do que as outras, como o Barroco com 12.60%, o Neogótico com 9.70%, e o Modernista com 11.70%. Esse desequilíbrio pode levar a um superajustamento e impactar o desempenho do modelo.

Tabela 5. Número de imagens dos conjuntos e *Imbalance Ratio* por estilo arquitetônico.

Estilo Arquitetônico	Imagens reais	Imagens sintéticas	Conjunto de treino total	Conjunto de treino	Proporção (%)
Barroco	260	0	260	65	12,60
Neoclássico	226	448	664	58	32,30
Neo-Gótico	200	0	200	50	9,70
Eclético	232	458	690	58	33,60
Modernista	240	0	240	60	11,70

3.6.5. Experimento 5: O uso do aumento de dados para todas as classes do conjunto de treinamento

O objetivo deste experimento é explorar o potencial do aumento de dados no treinamento do modelo de classificação de estilos arquitetônicos. Para isso, foram geradas imagens sintéticas por modelos de difusão e adicionadas a cada classe do conjunto de treinamento até que estejam equilibradas com 915 imagens, mantendo as imagens reais originais conforme detalhado na Tabela 6. Isso resultou em um conjunto de dados expandido para 4575 imagens, distribuídas igualmente entre todas as classes. Essa abordagem visa enriquecer a diversidade e representatividade do conjunto de treinamento para cada estilo arquitetônico, permitindo uma comparação com o modelo inicial de base e uma análise de como o aumento considerável do conjunto de treinamento influencia o desempenho do modelo.

Tabela 6. Número de imagens dos conjuntos por estilo arquitetônico.

Estilo Arquitetônico	Imagens reais	Imagens sintéticas	Conjunto de treino total	Conjunto de treino
Barroco	260	655	915	65
Neoclássico	226	689	915	58
Neo-Gótico	200	715	915	50
Eclético	232	683	915	58
Modernista	240	675	915	60

3.7 Divisão do conjunto de dados para treinamento do modelo

Neste estudo, foi priorizada a análise do desempenho do modelo com base nos dados de treinamento, adotando a proporção tradicional de 80% para o conjunto de treinamento e 20% para o conjunto de teste, observando sua evolução ao longo das épocas, utilizando métricas como acurácia e perda. Essa abordagem permitiu compreender como o modelo estava se adaptando aos dados e melhorando ao longo do tempo. Embora tenhamos focado nos dados de treinamento para o processo de ajuste do modelo, é importante destacar que o critério de parada do treinamento foi determinado pela avaliação do desempenho no conjunto de teste. Isso significa que o modelo foi refinado com base não apenas no seu desempenho nos dados de treinamento, mas também na sua capacidade de generalização para dados não vistos durante o treinamento. Essa metodologia proporcionou uma compreensão abrangente do processo de aprendizado do modelo, permitindo fazer ajustes estratégicos para melhorar sua performance e evitar o *overfitting*, que é uma preocupação comum em modelos de aprendizado de máquina. Embora não tenhamos formalmente separado conjuntos de validação e teste, essa abordagem permitiu compreender de maneira abrangente o processo de treinamento e refinamento do modelo, enfatizando a importância de monitorar o desempenho durante o treinamento, especialmente no conjunto de testes.

É importante ressaltar que o conjunto de testes permaneceu inalterado ao longo dos experimentos realizados. A variação ocorreu exclusivamente no conjunto de treinamento, que foi modificado em cada experimento. No Experimento 1, foi feita uma alteração no conjunto de dados de treinamento, onde foram inseridas imagens sintéticas, e para garantir o balanceamento, mantendo o volume de dados inalterado. Nos Experimentos 2 e 3, foram realizados aumentos de dados específicos para uma única classe em cada experimento. Essas alterações foram projetadas para avaliar o impacto do aumento de dados em classes individuais no desempenho geral do modelo. No Experimento 4, foram realizados aumento de dados nas duas classes mais problemáticas do modelo base, para avaliar a influência delas nas outras classes e na acurácia geral do modelo. Por outro lado, no Experimento 5, optamos por realizar um aumento de dados para todas as classes no conjunto de treinamento, buscando alcançar um melhor balanceamento entre as classes.

4 Análises, Avaliações e Resultados

4.1 Desempenho do modelo controle

No modelo treinado sem aumento de dados, para o estilo arquitetônico Barroco, observou-se uma taxa de acerto de 78,46% para conjunto de teste, com 51 imagens classificadas corretamente de um total de 65, enquanto para o estilo Eclético a taxa de acerto foi de 65,52%, com 38 imagens corretas de 58. Já para o estilo Modernistas, todas as 60 imagens foram corretamente classificadas, resultando em uma taxa de acerto de 100%. Para o estilo Neoclássico, a taxa de acerto foi de 60,34%, com 35 imagens corretas de 58, e para o estilo Neogótico, a taxa de acerto foi de 90,00%, com 45 imagens corretas de um total de 50, totalizando uma acurácia de modelo de 0,80350. A curva de aprendizagem do modelo pode ser vista nas Figuras 9 e 10, mostrando que não ocorreu superajustamento.

Podemos observar que as classes não possuem o mesmo número de imagens no conjunto de teste, com 65 instâncias para a classe Barroco, 58 para Eclético, 60 para o estilo Modernista, 58 para a classe Neoclássica e 50 instâncias para o estilo Neogótico, assim como o conjunto de treinamento, com 260, 232, 240, 226 e 200 instâncias para as classes Barroco, Eclético, Modernista, Neoclássico e Neogótico, respectivamente. Ao analisar as proporções de cada classe no conjunto de treinamento e no conjunto de teste, percebemos na Tabela 7 que não há um desequilíbrio significativo entre as classes, pois as proporções são semelhantes em ambos os conjuntos. Finalizando a análise desse modelo, é possível avaliar sua precisão e pontuação-F1 na Tabela 7.

Tabela 7. Acurácia e pontuações-F1 do modelo

Classe	Acurácia	Pontuação-F1	Conjunto de teste	Conjunto de treinamento
Barroco	78,46%	0,76	65	260
Eclético	65,52%	0,62	58	232
Modernista	100,00%	0,98	60	260
Neoclássico	60,34%	0,68	58	226
Neogótico	90,00%	0,90	50	200

Avaliar o desempenho de um modelo de classificação é essencial para compreender sua eficácia. A Matriz de Confusão, representada na Figura 11, oferece uma análise detalhada das previsões feitas pelo modelo em comparação com os valores reais, destacando os erros cometidos e as classificações mais precisas.

Analizando a matriz de confusão desse modelo, focamos nas duas classes com os menores valores de acurácia: os estilos Neoclássico e Eclético. Para o estilo Neoclássico, a maioria dos erros (13 de 23) ocorre quando o modelo o classifica como Eclético, 8 como Barroco e apenas 2 como Modernista. Já para o estilo Eclético, a maior parte dos erros (9 de 20) acontece quando o modelo o classifica como Barroco, 6 como Neoclássico e 5 como Neogótico. Além disso, observa-se uma taxa de acerto alta (100%) para o estilo modernista, e em segundo para o neogótico (90%).

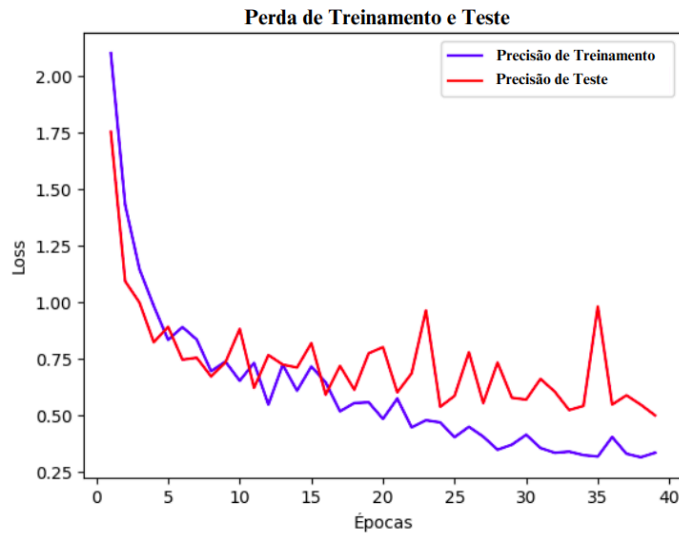


Fig. 9. Curva de loss para o conjunto de treino e teste

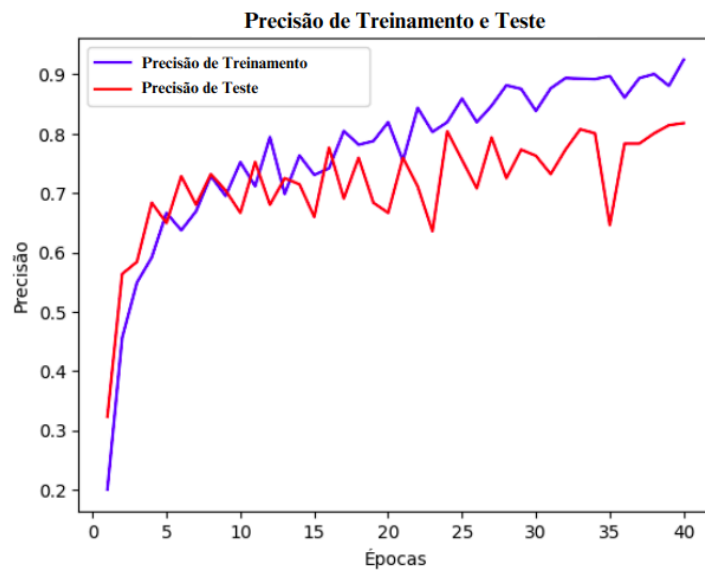


Fig. 10. Curva de precisão para o conjunto de treino e teste

Considerando apenas as métricas de precisão e acurácia até a época 40, observamos que o modelo alcançou resultados promissores. A precisão de treinamento atingiu aproximadamente 94,30%, enquanto a precisão no conjunto de teste foi de cerca de 80,76%. Da mesma forma, a acurácia de

treinamento foi em torno de 92,49%, e a acurácia do conjunto de teste chegou a aproximadamente 80,76%.

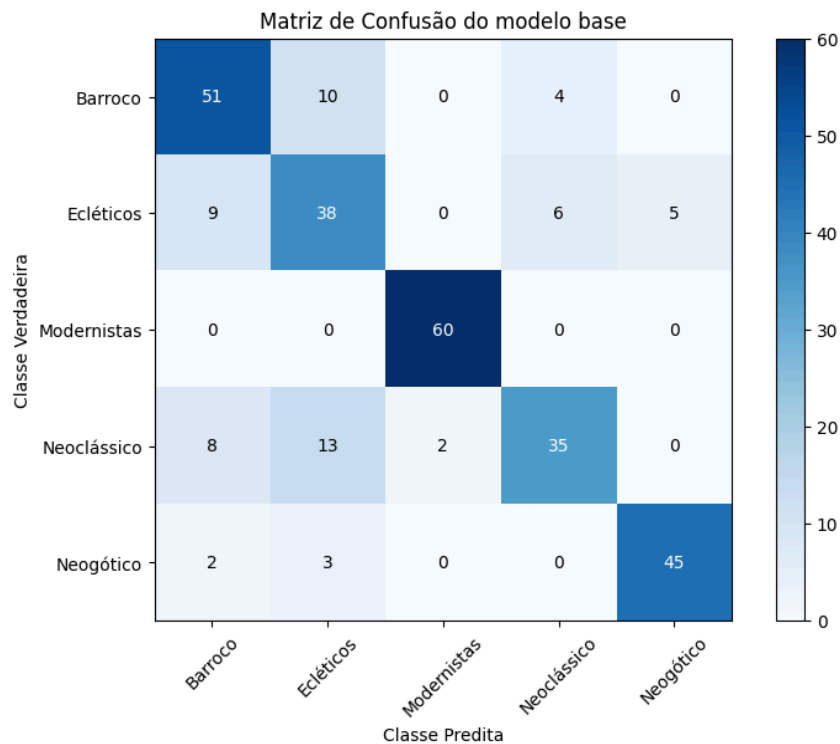


Fig. 11. Matriz de confusão do modelo base

4.2 Desempenho do modelo do Experimento 1

Com base nos gráficos fornecidos na Figura 12, o modelo mostrou uma melhoria gradual na acurácia e precisão ao longo das épocas, também é evidente que a acurácia do conjunto de teste não aumentou significativamente após um certo ponto, indicando uma possível estagnação. Além disso, podemos destacar que a acurácia no conjunto de teste não diminuiu significativamente em relação à acurácia de treinamento, sugerindo que o modelo não está superajustado aos dados de treinamento.

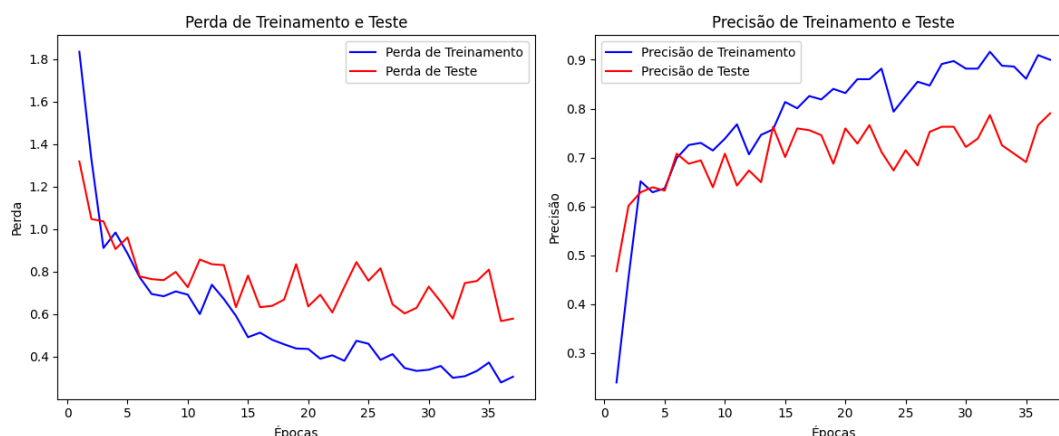


Fig. 12. Curva de perda e de precisão para o conjunto de treino e teste do experimento 1

Analisando a Tabela 8 e a Figura 13, referentes à eficiência do modelo e à matriz de confusão, respectivamente, percebe-se que o modelo se manteve estável em comparação ao modelo base. Houve uma pequena diminuição na assertividade na classe Barroco, de 51 para 49. Os outros classificadores permaneceram relativamente estáveis, com pequenas variações na assertividade, sendo a maior no classificador Eclético, com uma variação de 10%, de 0,65 para 0,75. Ao mesmo tempo em que houve uma melhora no desempenho do classificador Eclético, foi observada uma piora no desempenho do classificador Neoclássico, com uma diminuição de 3,44% em sua acurácia, onde imagens erroneamente classificadas como Barroco, passaram de 8 para 11.

Com base nos resultados obtidos, podemos inferir que o treinamento do novo modelo não sofreu grandes perdas, mesmo após substituir metade das imagens do conjunto de treinamento da classe Barroco por imagens sintéticas. Isso sugere que o uso de imagens sintéticas pode ter uma relevância significativa em projetos que lidam com a escassez de dados, porém não substituem com o mesmo valor as imagens reais. Essa constatação ressalta a capacidade das imagens sintéticas em complementar conjuntos de dados limitados.

Tabela 8. Acurácia e pontuação-F1 do Experimento 1

Classe	Acurácia modelo base	Acurácia	Pontuação - F1	Conjunto de treino	Conjunto de teste
Barroco	78,46%	75,38%	0,73	260	65
Eclético	65,52%	75,86%	0,72	232	58
Modernista	100,00%	100,00%	0,98	260	60
Neoclássico	60,34%	56,90%	0,65	226	58
Neogótico	90,00%	92,00%	0,92	200	50

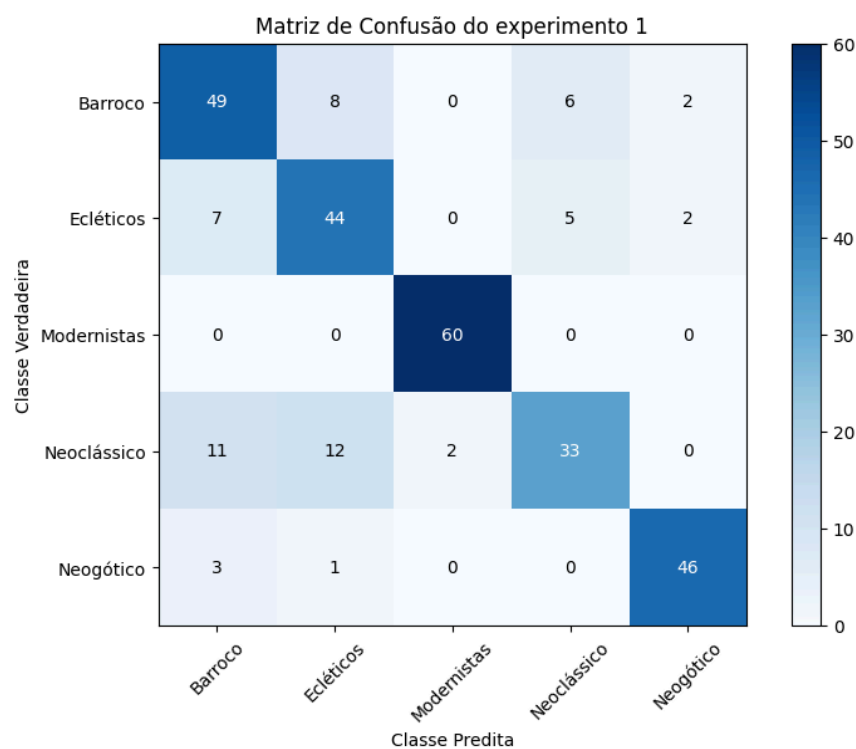


Fig. 13. Matriz de confusão experimento 1

4.3 Desempenho do modelo do Experimento 2

De acordo com a Figura 14, após 27 épocas de treinamento, o modelo atingiu seu melhor desempenho, alcançando uma precisão máxima de aproximadamente 91,7% no conjunto de treinamento e uma acurácia máxima de cerca de 79,4% no conjunto de teste, com uma perda de 0,550 neste último conjunto. A partir dessa época, não houve mais melhorias significativas na acurácia do conjunto de teste, estabilizando em torno de 0,7801, sugerindo que o modelo pode ter atingido seu limite de generalização para este conjunto de dados.

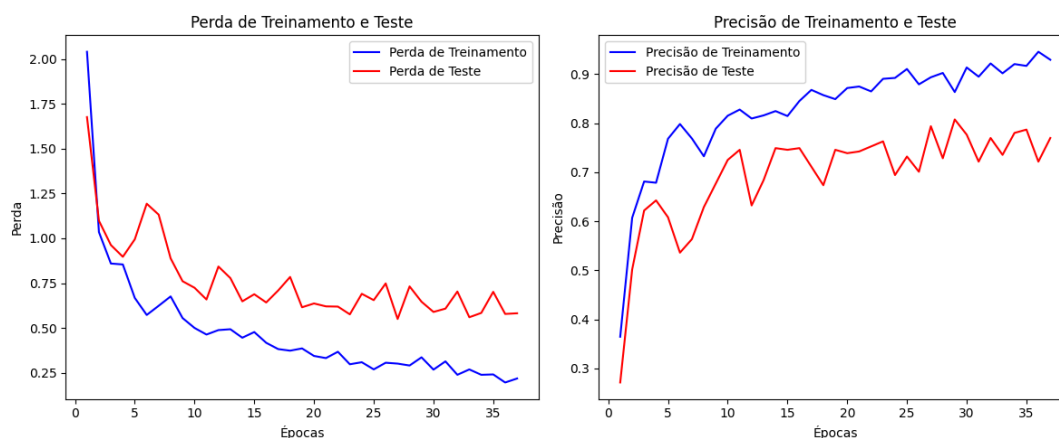


Fig. 14. Curva de perda e de precisão para o conjunto de treino e teste do Experimento 2.

Após a realização do aumento de dados para a classe neoclássica, observamos na Figura 15 uma melhoria na capacidade do modelo de classificar esse estilo arquitetônico. O aumento da taxa de acerto de 60,34% para 74,13% (como indicado na Tabela 9) sugere que a inclusão de mais dados permitiu que o modelo captasse com maior precisão as características distintas do neoclássico, resultando em uma redução de erros de classificação erroneamente atribuídos ao Barroco de 8 para 5, eclético de 13 para 9 e modernista de 2 para 1. A métrica de pontuação-F1 teve um aumento do modelo base para este, de 0,68 para 0,69.

No entanto, outros classificadores foram influenciados, passaram a apresentar maior número de erros de classificação neoclássica em seus conjuntos de teste. O classificador do estilo Barroco, por exemplo, passou de 4 para 11 erros, enquanto o do estilo Eclético aumentou de 6 para 10 erros e o Neogótico de 0 para 2, indicando possíveis desafios adicionais na classificação de outras classes após o ajuste.

Tabela 9. Acurácia e pontuações-F1 do Experimento 2

Classe	Acurácia modelo base	Acurácia	Pontuação- F1
Barroco	78,46%	66,15%,	0,73
Eclético	65,52%	74,15%	0,69
Modernista	100,00%	100,00%	0,98
Neoclássico	60,34%	74,14%	0,69
Neogótico	90,00%	84,00%	0,89

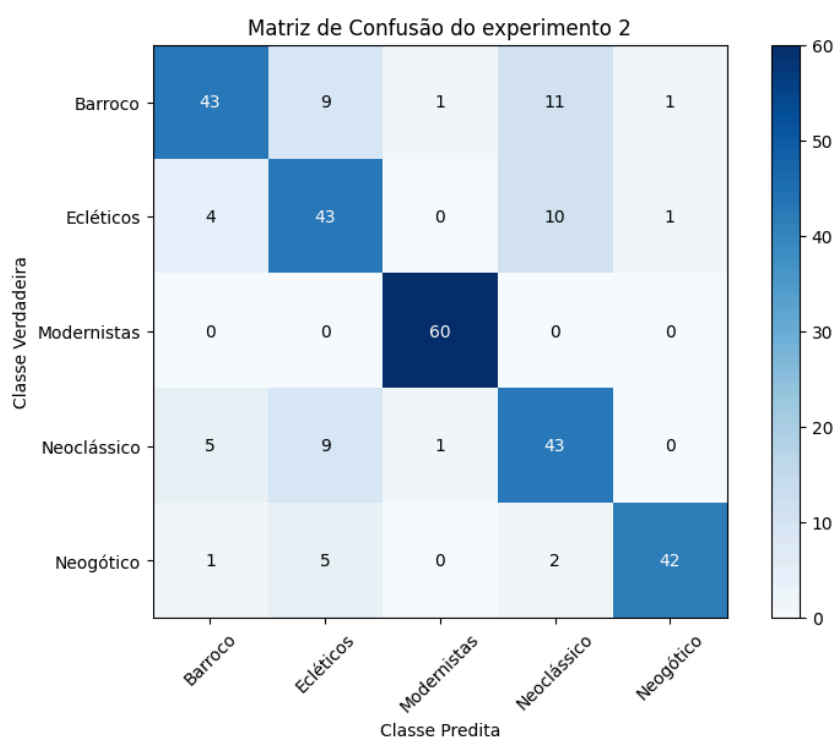


Fig. 15. Matriz de confusão experimento 2

4.3 Desempenho do modelo do Experimento 3

De acordo com os gráficos da Figura 16, é perceptível que durante as primeiras 50 épocas de treinamento do modelo, houve um progresso em termos de precisão e perda nos conjuntos de treinamento e teste. A precisão no conjunto de treinamento aumentou consistentemente de cerca de 36,63% para 95,11%, enquanto a função de perda diminuiu gradualmente. No conjunto de teste, a precisão acompanhou esse aumento, atingindo aproximadamente 80,70%. Vale ressaltar que o treinamento continuou além das 50 épocas planejadas, mas o modelo de melhor desempenho foi salvo de acordo com os critérios de parada estabelecidos na época 50. Esses resultados sugerem que o modelo está aprendendo efetivamente os padrões nos dados de treinamento e é capaz de generalizar bem para o conjunto de dados não vistos.

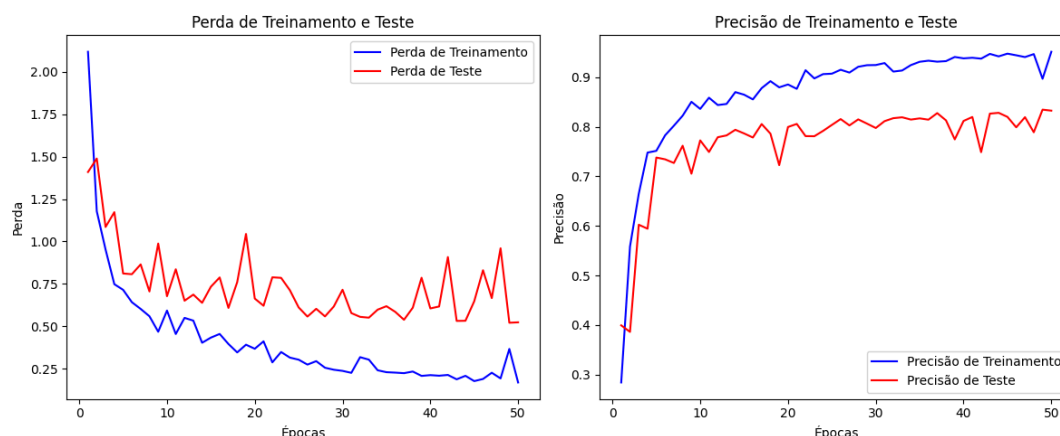


Fig. 16. Curva de perda e de precisão para o conjunto de treino e teste do Experimento 3.

Após a realização do aumento de dados para a classe eclética, foi observada, na Figura 17, uma significativa melhoria na capacidade do modelo em classificar esse estilo arquitetônico. O aumento da taxa de acerto de 65,52% para 77,59% (conforme indicado na Tabela 10) sugere que a inclusão de mais dados permitiu que o modelo captasse com maior precisão o conjunto de dados de teste da classe eclética, resultando em uma redução de erros de classificação erroneamente atribuídos ao Barroco de 9 para 4 e ao Neogótico de 5 para 2. A métrica de pontuação-F1 teve um aumento do modelo base para este, de 0,62 para 0,68.

No entanto, outros classificadores foram influenciados, passaram a apresentar maior número de erros de classificação neoclássica em seus conjuntos de teste. O classificador do estilo Neoclássico, por exemplo, passou de 13 para 14 erros, enquanto o do estilo Neogótico aumentou de 3 para 7 erros. Com exceção, do classificador Barroco que houve uma melhora de 10 para 8. Essas variações indicam possíveis desafios adicionais na classificação de outras classes após o ajuste.

Tabela 10. Precisão e pontuações-F1 do Experimento 3

Classe	Acurácia modelo base	Acurácia	Pontuação- F1
Barroco	78,46%	78,46%	0,82
Ecletico	65,52%	77,59%	0,68
Modernista	100,00%	100,00%	0,98
Neoclássico	60,34%	62,07%	0,67
Neogótico	90,00%	86,00%	0,90

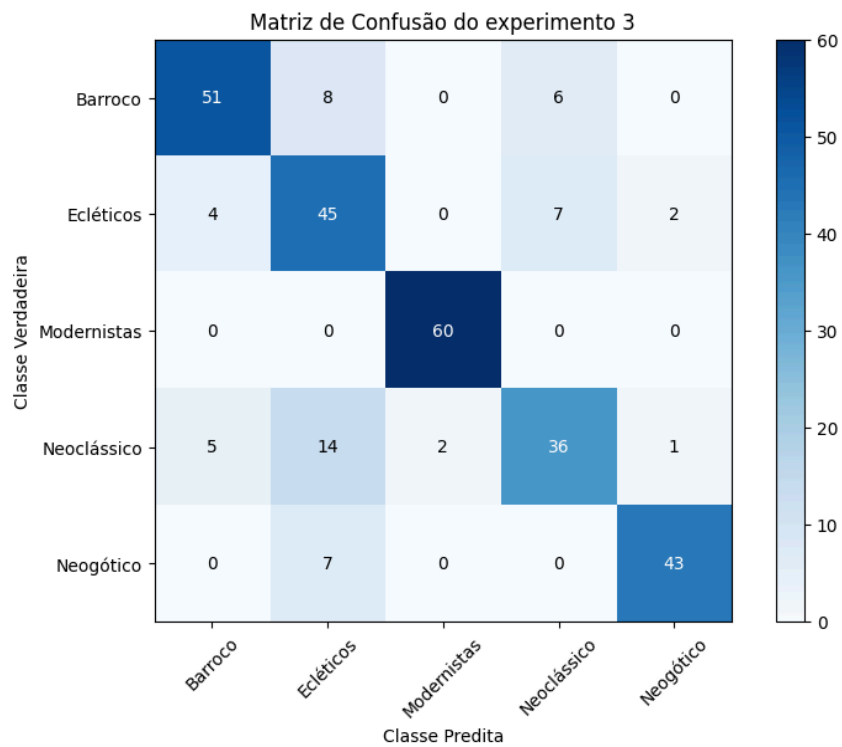


Fig. 17. Matriz de confusão Experimento 3.

4.4 Desempenho do modelo do Experimento 4

Durante o treinamento do modelo de classificação, observamos uma melhoria progressiva, tanto no desempenho do conjunto de treinamento, quanto no conjunto de teste ao longo das épocas. Inicialmente, a precisão de treinamento foi de aproximadamente 50%, enquanto a perda foi de 1,2064, enquanto no conjunto de teste a precisão foi de cerca de 63% com uma perda de 0,8883. Conforme o treinamento avançou, a precisão e a perda foram melhorando gradualmente para ambos os conjuntos. A melhor época, em termos de menor perda no conjunto de teste, foi a 20ª época, onde a perda foi de 0,5048 e a precisão de 83,51%.

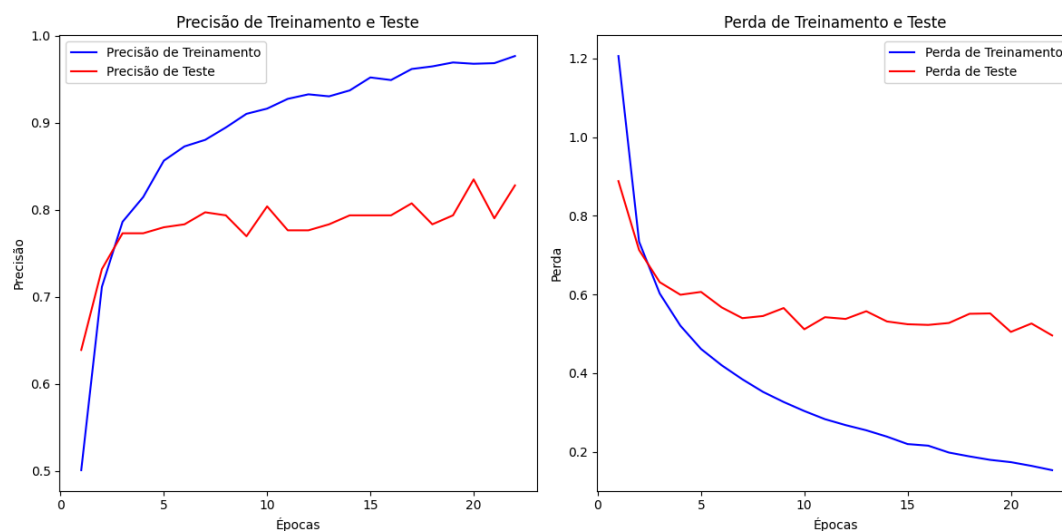


Fig. 18. Curva de perda e de precisão para o conjunto de treino e teste do Experimento 4.

Após a realização do aumento de dados para as classes neoclássica e eclética, foi observada, na Figura 19, uma significativa melhoria na capacidade do modelo em classificar esses estilos arquitetônicos. O aumento da taxa de acerto de 65,52% para 72,41% (conforme indicado na Tabela 10) para classe eclética e 60,34% para 77,51% sugere que a inclusão de mais dados permitiu que o modelo captasse com maior precisão do conjuntos de dados de teste das classes citadas. A métrica de pontuação-F1 da classe neoclássica, eclética e barroca, tiveram aumentos consideráveis para 0,85, 0,74, 0,71, respectivamente.

No entanto, outros classificadores foram influenciados no caso do estilo Barroco, passou a ter uma melhoria na sua acurácia, devido a redução das classificações ecléticas e neoclássicas diminuírem de 10 para 7 e 5 para 4, respectivamente. Já o classificador do estilo Neogótico apresentou uma pequena variação negativa na sua acurácia, de 90% para 88%, aumentando classificações errôneas de ecléticas e neoclássicas de 3 para 4 e 0 para 2, respectivamente. Essas variações indicam um sinal positivo, e nos diz respeito, a como a classe Barroco é influenciável com ajustes feitos as classes neoclássica e eclética.

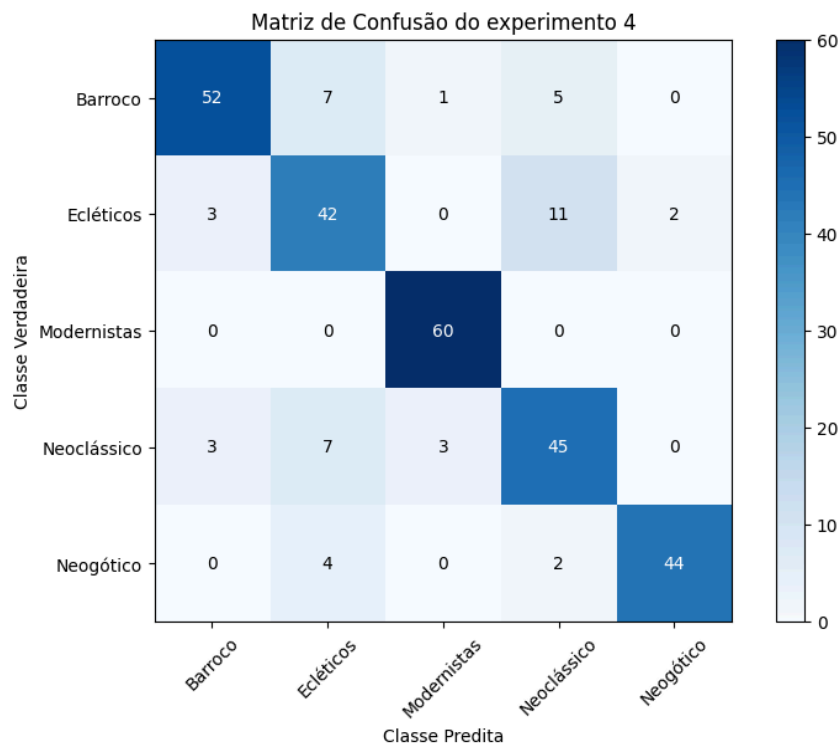


Fig. 19. Matriz de confusão Experimento 4

Tabela 11. Precisão e pontuações-F1 do Experimento 4

Classe	Acurácia modelo base	Acurácia	Pontuação-F1
Barroco	78,46%	80,00%	0,85
Eclético	65,52%	72,41%	0,71
Modernista	100,00%	100,00%	0,97
Neoclássico	60,34%	77,59%	0,74
Neogótico	90,00%	88,00%	0,92

4.5 Desempenho do modelo do Experimento 5

Durante o treinamento do modelo de classificação de imagens, observamos uma melhoria contínua, tanto no desempenho do conjunto de treinamento, quanto no conjunto de testes ao longo das épocas. Inicialmente, na primeira época, o modelo exibiu uma precisão de aproximadamente 67,06% no conjunto de treinamento, com uma perda de 0,8333, enquanto no conjunto de teste a precisão foi de cerca de 73,88%, com uma perda de 0,6382. Conforme o treinamento avançava, tanto a precisão quanto a perda foram gradualmente melhorando para ambos os conjuntos. O ponto

mais destacado ocorreu na 7ª época, onde a perda no conjunto de teste atingiu seu ponto mais baixo em 0,4747, acompanhada por uma precisão de 81,44%. Apesar de algumas flutuações ao longo das épocas, essa tendência de melhoria indica que o modelo estava aprendendo efetivamente os padrões nos dados de treinamento e generalizando bem para os dados de teste.

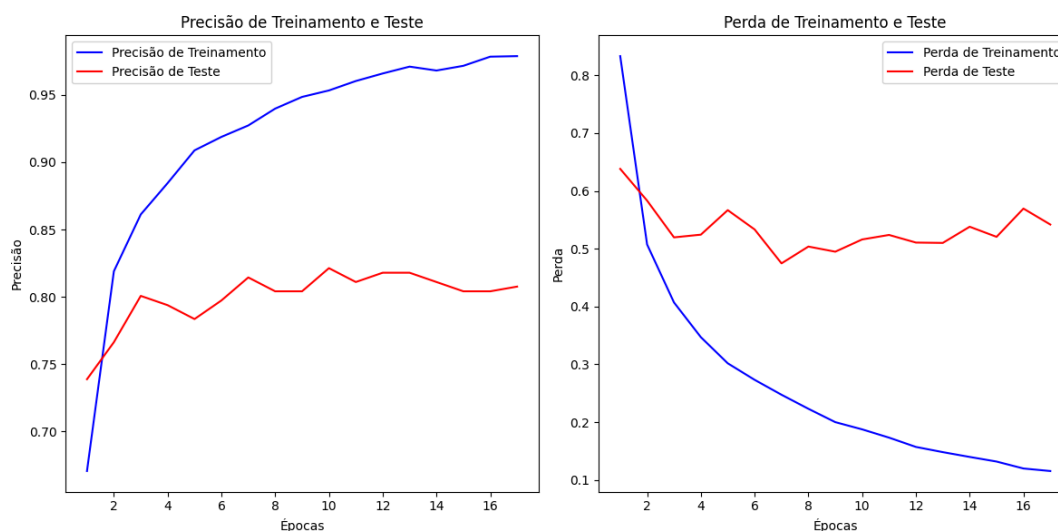


Fig. 20. Curva de perda e de precisão para o conjunto de treino e teste do Experimento 5.

Após a aplicação do aumento de dados em todas as classes, a análise na Figura 21 revelou que o desempenho do modelo permaneceu bastante semelhante ao do modelo base do experimento, com exceção da classe neoclássica. Nesta, foi observado um aumento notável na taxa de acerto, de 60,34% para 82,76% (como evidenciado na Tabela 12). Isso sugere que a inclusão de mais dados resultou em uma melhoria substancial no desempenho desta classe específica, enquanto as demais classes parecem ter alcançado uma estabilização em seus resultados.

A classe neoclássica se destacou com um desempenho superior durante o processo de classificação, evidenciado por resultados mais favoráveis, em comparação com outras classes. Em relação ao modelo base, observou-se uma redução significativa nas classificações incorretas, especificamente para as classes Barroco, Eclético e Modernista, onde as ocorrências diminuíram de 8 para 3, 13 para 6 e 2 para 1, respectivamente.

No entanto, ao examinarmos a matriz de confusão da Figura 21 e analisarmos os casos do conjunto de teste eclético, notamos uma tendência em que as imagens estão sendo frequentemente classificadas como neoclássicas. Inicialmente, houve 9 erros de classificação como barrocos, 6 como neoclássicos e 5 como neogóticos. Porém, após as modificações, os erros de classificação como barrocos reduziram para 2, enquanto os erros como neoclássicos aumentaram consideravelmente para 19, e os erros como neogóticos diminuíram de 5 para 2. Esses resultados sugerem que o modelo está mais bem treinado para identificar características neoclássicas do que características gerais do ecletismo, o que resulta em uma maior propensão a classificar imagens ecléticas como neoclássicas.

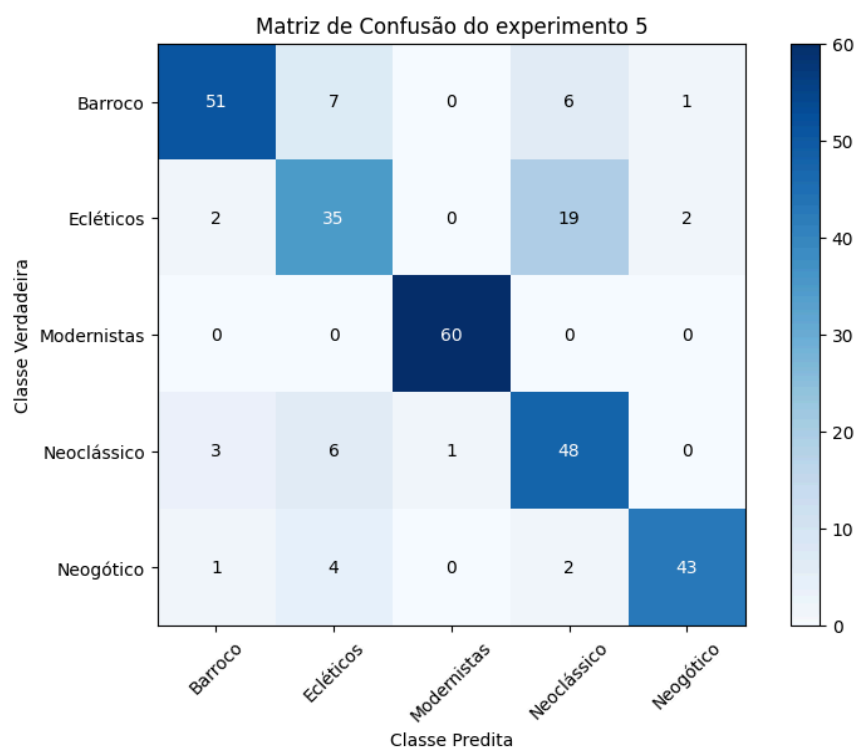


Fig. 21. Matriz de confusão Experimento 5.

Tabela 12. Precisão e pontuações-F1 do Experimento 5

Classe	Acurácia modelo base	Acurácia	Pontuação F1 modelo base	Pontuação- F1
Barroco	78,46%	78,46%	0,76	0,84
Eclético	65,52%	60,34%	0,62	0,64
Modernista	100,00%	100,00%	0,98	0,99
Neoclássico	60,34%	82,76%	0,68	0,72
Neogótico	90,00%	86,00%	0,90	0,90

4.5 Análise de Erros e Hipóteses de Causas

Nesta seção, aprofunda-se a compreensão do desempenho do modelo de classificação de estilos arquitetônicos, examinando casos em que o modelo fez previsões incorretas. O objetivo é identificar os motivos por trás desses erros, oferecendo entendimentos sobre as limitações do modelo em reconhecer e distinguir entre os diferentes estilos arquitetônicos. Ao explorar esses casos, busca-se não apenas compreender as falhas do modelo, mas também propor hipóteses sobre os fatores que podem influenciar sua capacidade de classificação. Com essa análise detalhada, espera-se destacar

áreas para melhorias futuras e aprimoramento do modelo. Começaremos examinando alguns exemplos de erros de classificação e discutindo as possíveis causas por trás deles.

4.5.1. O desafio do classificador do estilo arquitetônico Eclético

A complexidade em identificar o estilo eclético reside na sua capacidade de misturar uma variedade de elementos de diferentes estilos arquitetônicos, como por exemplo, neoclássico, neogótico e barroco, muitas vezes é difícil identificar claramente as características específicas de um monumento como sendo exclusivamente ecléticas. Como a classe eclética tem características muito fortes de outros estilos, é comum que o modelo reconheça a presença de elementos de diferentes estilos arquitetônicos, o que torna difícil decidir claramente classificá-la como eclética. Essa mistura de influências pode levar a interpretações diferentes e desafiar a atribuição de apenas uma classe para esse estilo arquitetônico, tornando o estilo eclético uma categoria mais desafiadora para definir com precisão.

A Figura 22 apresenta características muito evidentes de simetria e proporções clássicas, características distintivas da arquitetura neoclássica. O uso de colunas, uma das principais características desse estilo, contribui para a sensação de ordem e harmonia. Esses elementos são típicos de construções inspiradas na antiguidade clássica e são frequentemente associados ao neoclassicismo, logo induziu o modelo a interpretá-la como sendo do estilo Neoclássico.



Fig. 22. Monumento classificado como Neoclassico, pertencente à arquitetura Eclética.

A igreja da Figura 23 emana a essência do estilo arquitetônico barroco, construída, aparentemente, com materiais tradicionais, e apresenta linhas fortes e simetria perfeita. Sua fachada amarela se destaca, criando um contraste com o ambiente ao redor e evocando uma sensação de tempos passados. Esses elementos reforçam as características marcantes de uma igreja no estilo

barroco. Portanto, é compreensível que o modelo a classifique dessa forma, embora o criador da base de dados a tenha definido como estilo Eclético. O mesmo deve ter levado em consideração fatores históricos e de localização para a definição do estilo arquitetônico, fatores estes que não são levados em consideração no aprendizado de máquina.



Fig. 23. Monumento classificado como Barroco, pertencente ao conjunto de teste da classe Eclética.

O monumento da Figura 24 possui simetria e colunas clássicas que remetem ao neoclassicismo, assim como detalhes esculpidos que lembram o estilo barroco, fachadas e ornamentações bem elaboradas, e uma proporção grandiosa em sua construção, ressaltando características mais ecléticas do que neoclássicas. No entanto, é possível perceber uma influência greco-romana em sua construção, característica marcante do estilo neoclássico e bastante presente no conjunto de treinamento. Sendo assim, o peso dessas características induz o modelo a classificar erroneamente a estrutura como pertencente ao conjunto de monumentos do estilo arquitetônico neoclássico. Outro ponto importante a considerar sobre o erro é o impacto da fotografia do monumento durante a noite. Dado que o número de imagens no conjunto de treinamento tiradas à noite é significativamente menor, pode haver uma escassez de representações adequadas dos detalhes arquitetônicos importantes. Esses detalhes, mais perceptíveis em uma imagem mais clara, podem dificultar a análise precisa e a classificação correta do estilo arquitetônico do monumento.

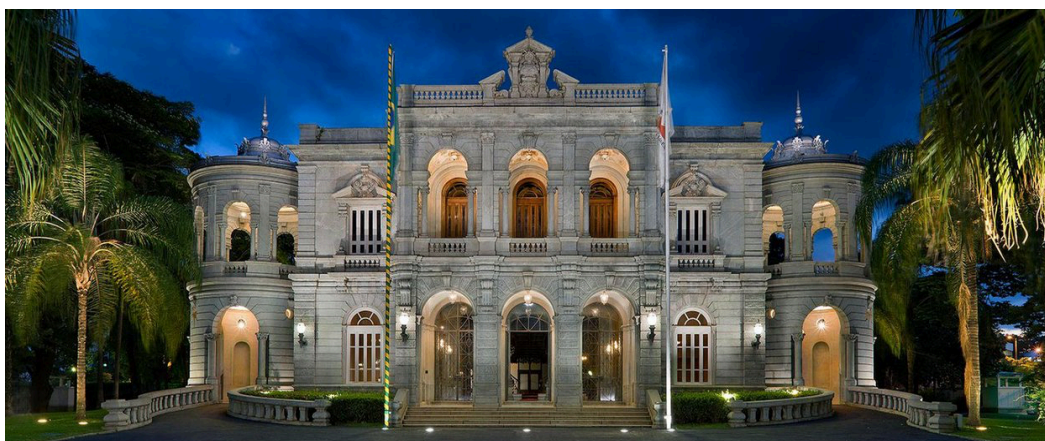


Fig. 24. Monumento classificado como Neoclássico, pertencente ao conjunto de teste da classe Eclética.

4.5.2. Classificações erradas no conjunto de teste Neoclássico

Na Figura 25, é possível observar expressões características tanto do neoclassicismo, quanto do estilo barroco. O modelo enfrentou dificuldades para distinguir entre esses dois estilos arquitetônicos, já que estavam muito próximos em termos de pontuação. No entanto, o barroco se destacou com a maior pontuação. Isso sugere que os detalhes ornamentais nas janelas e na porta principal e seus traços azuis de linha forte, típicos do barroco, tiveram mais peso para o modelo do que as características neoclássicas, como a simetria, linhas retas e o formato triangular no topo central do monumento.



Fig. 25. Monumento classificado como Barroco, pertencente ao conjunto de teste da classe Neoclássico.

A Figura 26 retrata um monumento que combina fortes características do neoclassicismo e do ecletismo. O neoclassicismo é um movimento artístico que valoriza a inspiração nas formas arquitetônicas da Grécia e Roma antigas, enquanto rejeita os excessos ornamentais, encontrados por exemplo, no Barroco. Tais características são fortemente presentes neste monumento, como colunas imponentes, similares às estruturas grego-romanas, e fachada simétrica.

Quanto ao ecletismo, este surge da combinação de diferentes influências arquitetônicas, resultando em monumentos que mesclam elementos de estilos variados, permitindo liberdade criativa e uma diversidade de detalhes ornamentais. Na Figura 26, a presença de detalhes ornamentais nas molduras das janelas e na entrada principal, a simetria que também não se limita ao neoclassicismo, a sensação que a obra passa de grandiosidade monumental são indícios de um estilo eclético presente. Também é possível notar o uso de materiais como vidro com aparência laminada. Essas características sugerem uma diversidade de influências arquitetônicas. Todas essas características parecem ter sido mais importantes para o modelo na hora de decidir, superando as características do estilo puramente neoclássico. Assim, justificando a decisão do modelo de classificar este monumento como eclético.



Fig. 26. Monumento classificado como Eclético, pertencente ao conjunto de teste da classe Neoclássico.

5 Considerações Finais e Trabalhos Futuros

Através deste estudo, foi possível compreender a importância e o potencial do uso de técnicas mais avançadas de aumento de dados, como os modelos de difusão latente, na melhoria da

performance dos modelos de classificação. Levando-se em consideração bases de dados limitadas, sua aplicabilidade torna-se de grande importância por mitigar os efeitos da escassez de dados. Além disso, ao utilizar técnicas de modelos de difusão para a geração de novas imagens, conseguimos explorar regiões mais abrangentes e complexas do espaço de características, permitindo uma melhor representação das classes e uma adaptação mais eficaz a exemplos não vistos anteriormente. Por vezes, os erros de classificação deste estudo podem ser atribuídos à similaridade visual entre certos estilos arquitetônicos, que compartilham características ornamentais e estruturais similares.

É fundamental reconhecer as limitações inerentes aos modelos de difusão, especialmente em relação à eficiência computacional durante a inferência. A geração de cada imagem adicional demanda uma quantidade significativa de tempo e recursos computacionais, devido à complexidade das etapas de avaliação envolvidas. Esse processo pode ser demorado, tornando desafiadora a escalabilidade e a aplicação prática desses modelos em conjuntos de dados de grande volume. O modelo *Stable Diffusion* apresenta suas próprias limitações importantes que merecem atenção. Por exemplo, a qualidade das imagens geradas pode diminuir consideravelmente quando utilizamos imagens de entrada com resoluções menores que 450x450. Além disso, surgem dificuldades na geração de imagens precisas, quando a resolução da imagem é relativamente baixa. Adaptar o modelo para tarefas específicas exige uma quantidade considerável de dados e treinamento, sendo sensível à qualidade desses novos dados. Dessa forma, é importante ressaltar que o sucesso dessas técnicas está ligado à qualidade dos dados sintéticos gerados e à minuciosa avaliação de seus impactos na performance dos modelos. É crucial considerar o conhecimento prévio das características do classificador para o qual as novas imagens estão sendo geradas, destacando a importância de um entendimento profundo do modelo e de seu espaço de características.

A geração de imagens sintéticas nos Experimentos 2, 3 e 4 possibilitou algumas melhorias nos classificadores previstos, demonstrando que o uso da técnica de aumento de dados por meio de modelos de difusão é capaz de contribuir para a performance do modelo dentro de suas limitações, embora nos Experimentos 2 e 3 também foi percebido que o aumento de dados em uma única classe, tende interferir negativamente em algumas outras classes do modelo. No entanto, no Experimento 1, é importante perceber nos resultados que o uso de imagens reais tende a ter um peso maior, do que de imagens sintéticas para o modelo. Portanto, a aplicação dessa técnica para o aumento de dados não apenas impulsiona o desempenho dos modelos de classificação, mas também se configura como uma abordagem promissora para superar os obstáculos originados pela falta de dados em diferentes domínios de aplicação.

Para trabalhos futuros, uma abordagem interessante seria explorar ainda mais as capacidades do aumento de dados por meio do *stable diffusion*, utilizando a versão mais atual e disponível desse método. Além de modificar o monumento arquitetônico em si, como já realizado neste estudo, poderíamos considerar a variação do *background*, incorporando mudanças de fases do ano, como diferentes estações, e momentos do dia, como amanhecer, tarde e noite. Isso permitiria uma maior diversificação das imagens geradas, aumentando a robustez do conjunto de dados. Além disso, seria válido investigar a combinação do aumento de dados via *stable diffusion*, com técnicas mais convencionais de aumento de dados, como *flip*, rotação, deslocamento e zoom. Essa abordagem híbrida poderia potencializar os benefícios de ambas as técnicas, fornecendo uma base de treinamento ainda mais variada e representativa. Essas direções de pesquisa têm o potencial de melhorar significativamente o desempenho e a generalização dos modelos de classificação de estilos

arquitetônicos, expandindo suas capacidades, a fim de melhorar acurácia do modelo de maneira geral.

Para assegurar a confiabilidade dos modelos desenvolvidos, adotou-se uma metodologia de divisão dos conjuntos de dados em treino e teste. O conjunto de treino foi crucial para ensinar o modelo sobre os padrões arquitetônicos presentes nas imagens, enquanto o conjunto de teste foi empregado tanto no processo de treinamento, utilizando o método de parada, quanto na avaliação final do desempenho do modelo. No entanto, visando uma validação mais rigorosa e precisa, reconheceu-se a importância de criar um conjunto de validação separado. Este conjunto de validação seria utilizado exclusivamente para ajustar os hiperparâmetros do modelo, evitando o overfitting e garantindo uma generalização robusta para novos dados. Assim, o conjunto de teste permaneceria desconhecido para o modelo até a fase final de avaliação, garantindo uma análise imparcial do desempenho geral do modelo.

Outro passo interessante rumo a uma melhor acurácia do modelo, que não diz respeito mais a geração de imagens sintéticas, seria explorar como incluir informações contextuais, como ano de construção e localização geográfica dos monumentos, diretamente durante o treinamento do modelo. Isso poderia ser realizado criando características extras que representam esses dados contextuais e as incorporando como parte das entradas da rede neural. Essa combinação permitiria que o modelo aprendesse a identificar padrões tanto nas características visuais dos monumentos quanto nos dados contextuais, o que poderia resultar em uma melhor compreensão do contexto histórico e geográfico, e, consequentemente, uma classificação mais precisa dos estilos arquitetônicos.

Referências

1. ALPAYDIN, Ethem. Introduction to machine learning. MIT press, 2020.
2. PRAMOD, Akshara; NAICKER, Harsh Sankar; TYAGI, Amit Kumar. Machine learning and deep learning: Open issues and future research directions for the next 10 years. Computational analysis and deep learning for medical care: Principles, methods, and applications, p. 463-490, 2021.
3. AKROUT, Mohamed et al. Diffusion-based data augmentation for skin disease classification: Impact across original medical datasets to fully synthetic images. In: International Conference on Medical Image Computing and Computer-Assisted Intervention. Cham: Springer Nature Switzerland, p. 99-109, 2023.
4. MORENO-BAREA, Francisco J.; JEREZ, José M.; FRANCO, Leonardo. Improving classification accuracy using data augmentation on small data sets. Expert Systems with Applications, v. 161, p. 113696, 2020.
5. NGASA, Emmanuel Edward et al. Diffusion-based Wasserstein generative adversarial network for blood cell image augmentation. Engineering Applications of Artificial Intelligence, v. 133, p. 108221, 2024.
6. GARCIA, Ashley; LOPEZ, Melissa. Enhancing Machine Learning in Medical Imaging: Exploring Data Augmentation and Generative Models.
7. WANG, Xiaolong; HE, Zhijian; PENG, Xiaojian. Artificial-Intelligence-Generated Content with Diffusion Models: A Literature Review. Mathematics, v. 12, p. 977, 2024.
8. CROITORU, Florinel-Alin et al. Diffusion models in vision: A survey. IEEE Transactions on Pattern Analysis and Machine Intelligence, 2023.

9. ALHABEED, Sarah K.; AL-SHARGABI, Amal A. Text-to-Image Synthesis with Generative Models: Methods, Datasets, Performance Metrics, Challenges, and Future Direction. *IEEE Access*, 2024.
10. ROMBACH, Robin et al. High-resolution image synthesis with latent diffusion models. In: *Proceedings of the IEEE/CVF conference on computer vision and pattern recognition*. p. 10684-10695, 2022.
11. ZHANG, Tianyi et al. A Survey of Diffusion Based Image Generation Models: Issues and Their Solutions. *arXiv preprint arXiv:2308.13142*, 2023.
12. SUKKAR, A. W. et al. Analytical Evaluation of Midjourney Architectural Virtual Lab: Defining Major Current Limits in AI-Generated Representations of Islamic Architectural Heritage. *Buildings*, v. 14, p. 786, 2024.
13. ZHOU, Tong et al. Generative AI Enables EEG Data Augmentation for Alzheimer's Disease Detection Via Diffusion Model. In: *2023 IEEE International Symposium on Product Compliance Engineering-Asia (ISPCE-ASIA)*. IEEE, p. 1-6, 2023.
14. MORENO, Hugo et al. Analysis of Stable Diffusion-derived fake weeds performance for training Convolutional Neural Networks. *Computers and Electronics in Agriculture*, v. 214, p. 108324, 2023.
15. BORJI, Ali. Generated faces in the wild: Quantitative comparison of stable diffusion, midjourney and dall-e 2. *arXiv preprint arXiv:2210.00586*, 2022.
16. STÖCKL, Andreas. Evaluating a synthetic image dataset generated with stable diffusion. In: *International Congress on Information and Communication Technology*. Singapore: Springer Nature Singapore, 2023. p. 805-818.
17. DIAS, Gustavo Henrique Maia. Classificação de Estilo Arquitetônico através de Rede Neural Convolutacional Potencializada por Unidades de Processamento Neural para Dispositivos Móveis
18. SHAHOVEISI, F.; TAHERI GORJI, H.; SHAHABI, S.; HOSSEINIRAD, S.; MARKELL, S.; VASEFI, F. Application of Image Processing and Transfer Learning for the Detection of Rust Disease. *Scientific Reports*, Nature Publishing Group UK London, vol. 13, pp. 5133, 2023.
19. CERÁVOLO, Ana Lúcia. Interpretações do patrimônio: arquitetura e urbanismo moderno na constituição de uma cultura de intervenção no Brasil, anos 1930-60. 2010. Tese de Doutorado. Universidade de São Paulo.
20. BORAWAR, Lokesh; KAUR, Ravinder. ResNet: Solving vanishing gradient in deep networks. In: *Proceedings of International Conference on Recent Trends in Computing: ICRTC 2022*. Singapore: Springer Nature Singapore, p. 235-247, 2023.
21. KATTENBORN, Teja et al. Review on Convolutional Neural Networks (CNN) in vegetation remote sensing. *ISPRS journal of photogrammetry and remote sensing*, v. 173, p. 24-49, 2021.
22. BHATT, D. et al. CNN Variants for Computer Vision: History, Architecture, Application, Challenges and Future Scope. *Electronics*, v. 10, p. 2470, 2021.
23. SONG, Jia et al. A survey of remote sensing image classification based on CNNs. *Big earth data*, v. 3, p. 232-254, 2019.
24. AMJOUR, Ayoub Benali; AMROUCH, Mustapha. Object detection using deep learning, CNNs and vision transformers: a review. *IEEE Access*, 2023.
25. POOLA, Rahul Gowtham et al. COVID-19 diagnosis: A comprehensive review of pre-trained deep learning models based on feature extraction algorithms. *Results in Engineering*, p. 101020, 2023.
26. AKHTAR, Nadeem; RAGAVENDRAN, U. Interpretation of intelligence in CNN-pooling processes: a methodological survey. *Neural computing and applications*, v. 32, p. 879-898, 2020.

27. ALKHOULY, Asmaa A.; MOHAMMED, Ammar; HEFNY, Hesham A. Improving the performance of deep neural networks using two proposed activation functions. *IEEE Access*, v. 9, p. 82249-82271, 2021.
28. DING, Ning et al. Parameter-efficient fine-tuning of large-scale pre-trained language models. *Nature Machine Intelligence*, v. 5, p. 220-235, 2023.
29. LIPTON, Zachary Chase; ELKAN, Charles; NARAYANASWAMY, Balakrishnan. Thresholding classifiers to maximize F1 score. *arXiv preprint arXiv:1402.1892*, 2014.