

Applying Diffusion Models for Classification of Central Segregation in Steel Plates

Luciano M. Oliveira¹, Guilherme Silva¹, Arthur Negrão¹,
Andre L. Ottoni², Pedro Silva²

¹ Postgraduate Program in Computer Science – Federal University of Ouro Preto (UFOP)
35400-000 – Ouro Preto – MG – Brazil

²Computing Department – Federal University of Ouro Preto (UFOP)
35400-000 – Ouro Preto – MG – Brazil

{luciano.menezes, guilherme.lopes, arthur.negrao}@aluno.ufop.edu.br
{andre.ottoni, silvap}@ufop.edu.br

Abstract. *The automatic classification of segregation levels in continuous casting slabs is crucial for ensuring steel quality. Macrosegregation, characterized by localized chemical variations, poses a significant challenge, especially in the central region of the slabs. This work proposes an approach based on convolutional neural networks, specifically EfficientNet-B0 with transfer learning, combined with the generation of synthetic data through diffusion models (Stable Diffusion). Experimental results obtained on a dataset of 150 training images and 30 test images indicated an initial accuracy of 80% without any data augmentation, with an improvement of over 13% after the inclusion of synthetic data, as well as gains in precision and recall. The proposed method contributes to the automation of defect inspection and aligns with the principles of Industry 4.0.*

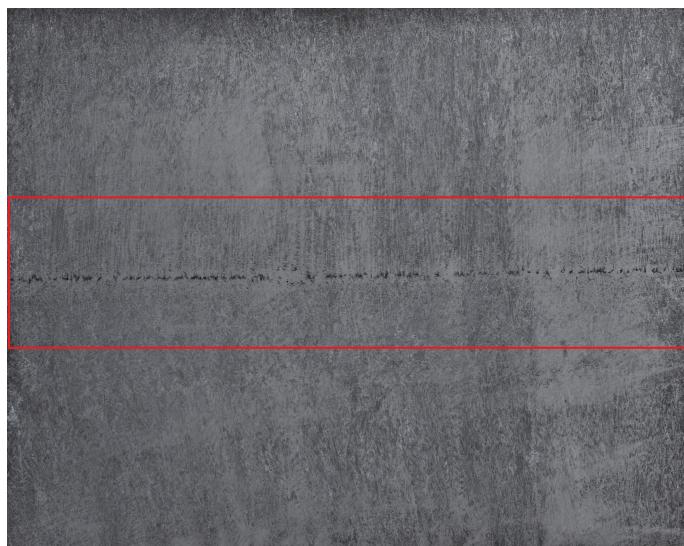
Resumo. *A classificação automática dos níveis de segregação em placas de lingotamento contínuo é crucial para garantir a qualidade do aço. A macrossegregação, caracterizada por variações químicas localizadas, representa um desafio significativo, especialmente na região central das placas. Este trabalho propõe uma abordagem baseada em redes neurais convolucionais, especificamente a EfficientNet-B0, com transferência de aprendizado, aliada à geração de dados sintéticos por meio de modelos de difusão (Stable Diffusion). Os resultados experimentais obtidos em um conjunto de treino de 150 imagens para treino e 30 para teste indicaram acurácia inicial de 80% sem nenhum aumento de dados, com uma melhora superior a 13% após a inclusão de dados sintéticos, além de ganhos em precisão e revocação. A proposta contribui para a automação da inspeção de falhas e se alinha aos princípios da Indústria 4.0.*

1. Introdução

A demanda por processos industriais mais eficientes e precisos tem impulsionado a Indústria 4.0, marcada pela integração de tecnologias digitais [Lima and Gomes 2020]. Na siderurgia, a automação de laboratórios é fundamental para garantir a precisão e confiabilidade das análises de propriedades do aço, assegurando a qualidade. Um dos principais desafios é a classificação da segregação central em placas de lingotamento contínuo, fenômeno metalúrgico complexo que afeta a qualidade do produto final [Ludwig et al. 2015].

A segregação central ocorre devido à concentração de elementos químicos como carbono e manganês durante a solidificação, impactando a integridade estrutural das placas [Moore 1984, Beckermann 2002]. Sua identificação geralmente depende de inspeção visual de amostras atacadas quimicamente de um especialista, um método subjetivo e demorado. A Figura 1 mostra o local típico de ocorrência da segregação central.

Figura 1. Local de presença da segregação central após retirada da amostra e imagem.



Nem sempre há uma disponibilidade alta de dados para treinamento de modelos de visão computacional que já são úteis para a identificação de padrões metalográficos [Queiroz 2015]. Contudo, há estratégias na literatura para a geração de imagens sintéticas para diversos contextos [Costa et al. 2024].

Considerando a limitação de dados reais, este estudo também investiga o uso de imagens sintéticas geradas por modelos de difusão que têm como objetivo aumentar os dados de treinamento do modelo de rede neural para melhorar a acurácia dos testes. Este estudo contribui para a melhoria da acurácia de redes neurais convolucionais na classificação da segregação central em placas de aço através da combinação de modelos de difusão e tipos de variação de dados abordados em [Shorten and Khoshgoftaar 2019] para confirmar esse objetivo.

Os resultados indicam uma acurácia inicial nos testes de 80,01%, com separação adequada entre as classes. A aplicação de CNNs demonstrou potencial para automatizar a classificação da segregação, reduzindo a subjetividade da análise humana. Após a geração de imagens sintéticas de cada nível e treinamento do modelo em imagens reais e sintéticas, os resultados de acurácia subiram para 86,67% e, após implementar *data augmentation* aos dados de treinamento, foi obtido uma acurácia de 93,33%.

Este trabalho está organizado da seguinte maneira: a Seção 2 apresenta a revisão da literatura; a Seção 3 apresenta a metodologia utilizada; já a Seção 4 apresenta os resultados obtidos; e, por fim, a Seção 5 apresenta a conclusão desta pesquisa e sugere direcionamentos para trabalhos futuros.

2. Trabalhos Relacionados

A aplicação de CNNs na indústria metalúrgica tem sido amplamente estudada, desde a detecção de defeitos superficiais até a previsão de propriedades mecânicas do aço [Gao et al. 2019], esses avanços têm permitido o aprimoramento dos processos industriais, otimizando inspeções e aumentando a eficiência do controle de qualidade. Estudos como o de [Zhang and Yang 2023] abordam como a aplicação de técnicas de *machine learning* estão presentes na indústria, revisando o estado da arte em processos siderúrgicos da Aciaria.

Em [Queiroz 2015], os autores realizam estudos fundamentais sobre a detecção da região de segregação de imagens. O método Canny, que apresentou os melhores resultados no trabalho citado, detecta bordas equilibrando boa localização e baixo ruído, usando filtros, cálculo de gradiente e dupla limiarização. A previsão da segregação é outro problema solucionado por meio de dados de processo apresentado no estudo [Nieto et al. 2018], ele utiliza *Support Vector Machines* otimizado com *Particle Swarm Optimization* superando uma rede neural perceptron multicamadas na comparação. O artigo [Zou et al. 2019] também realiza essa previsão da segregação em tarugos de lingotamento contínuo a partir de dados do processo, utilizando uma rede neural rasa. Ao utilizar modelos de regressão linear múltipla o resultado estava 57%, ao utilizar *Regularized Extreme Learning Machine*(R-ELM) os resultados obtidos foram maiores que 94%.

O artigo [Hu et al. 2024] investiga a identificação do estado da chama na Aciaria, mostrando a eficiência de redes neurais na classificação de imagens, especialmente a EL-DenseNet, as métricas utilizadas também foram eficiência, precisão, matriz de confusão e revocação. Já [Fu et al. 2023] utilizam CNNs simples, como SqueezeNet com múltiplas escalas para a classificação de defeitos superficiais, aumentando a precisão das análises. A precisão neste último trabalho teve um bom aumento, saindo de 88% com métodos tradicionais e superando 94% ao utilizar CNNs.

Em [Ibrahim 2024] é explorada a transferência de aprendizado para classificar padrões de defeitos de superfície em aços, demonstrando que redes neurais pré-treinadas podem ser eficazes e adaptadas para esse tipo de análise. O estudo utiliza a arquitetura VGG16 e uma nova CNN para classificar seis tipos de defeitos superficiais, em um conjunto público de imagens de aços laminados a quente. Os resultados obtidos foram promissores, com o modelo atingindo 99,44% de acurácia e métricas de precisão superiores a 98% na maioria das classes. No entanto, os autores reportaram que a arquitetura proposta apresenta alto consumo de memória.

Embora os trabalhos revisados representem avanços na aplicação de inteligência artificial à metalurgia, observa-se uma predominância do uso de arquiteturas convencionais de redes neurais convolucionais ou redes rasas, com pouca atenção dedicada a técnicas mais sofisticadas de generalização e aumento de dados. Em contraste, a abordagem proposta neste trabalho integra CNN, especificamente a EfficientNet B0 que é uma arquitetura leve e eficiente [Tan and Le 2019], com modelos de difusão [Ruiz et al. 2022] para *data augmentation*, o que permite não apenas aprimorar a robustez do modelo, mas também expandir de forma sintética e controlada a variabilidade do conjunto de treinamento.

Trabalhos que utilizam o aumento de dados de difusão como [Tai et al. 2021], utilizam tecnologias diferentes das utilizadas neste trabalho. É utilizado *Textual Inversion*,

que não realiza o *fine-tuning* assim como o *Dreambooth*, mas *Stable Surface Defect Generation* utiliza *Low-Rank Adaptation* e teve um aumento na acurácia após essa utilização entre 12% até 20%. O artigo [Li et al. 2025] utiliza uma abordagem parecida com *Stable Industrial Defect Generation* também com base em *Stable Diffusion* mas realizando o *fine-tuning* com *Low-Rank Adaptation* e também aumentou a acurácia no mesmo percentual relatado.

Outro trabalho [Zhang et al. 2024] que utiliza modelos de difusão como *Stable Diffusion* e *Dreambooth*, assim como este presente trabalho, utiliza *Denoising Diffusion Probabilistic Model* (DDPM), para prever a evolução microestrutural de materiais a partir de descrições textuais. O trabalho mostra que os valores de métricas como diâmetro médio de grãos e espaçamento dendrítico ficaram muito próximos dos valores reais.

Em [Ottoni et al. 2023] aborda o aumento de dados de forma tradicional aplicando transformações geométricas como zoom, deslocamentos e *flips* em imagens de construções civis para classificação com diferentes redes neurais convolucionais. O artigo propõe uma metodologia de ajuste de hiperparâmetros utilizando regressão logística. Os autores analisaram combinações de transformações, identificando um conjunto ótimo que elevou a acurácia de classificação para até 95,6% no caso de reconhecimento de vegetação em fachadas, evidenciando a importância do ajuste cuidadoso de *data augmentation* para aplicações com dados limitados.

3. Metodologia

Nesta seção, será mostrada a metodologia utilizada para a aquisição, criação, processamento e classificação das imagens macrográficas das amostras de aço. Como dito na Seção 1, o objetivo é identificar e classificar os níveis de segregação central em placas de lingotamento contínuo.

3.1. Preparação das Amostras e Aquisição da Base de Dados

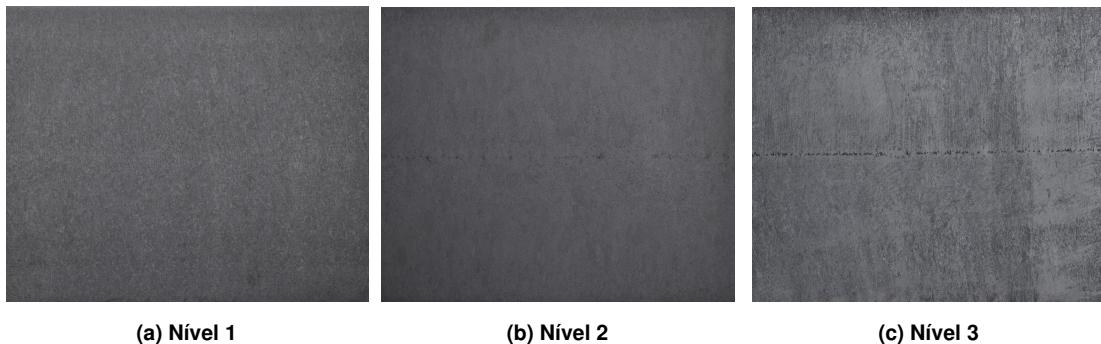
A preparação das amostras foi realizada conforme as práticas padronizadas de análise macrográfica. O objetivo desse processo é revelar a macroestrutura da amostra, destacando a região segregada e demais descontinuidades inerentes ao processo de lingotamento contínuo. O procedimento seguiu as seguintes etapas:

- Corte da amostra: As amostras foram cortadas em seções representativas da face frontal de placas do lingotamento contínuo.
- Preparação: A superfície das amostras foi lixada para remover impurezas e preparar a superfície para o ataque químico.
- Ataque químico: As amostras foram imersas em uma solução ácida específica, que promove a corrosão da superfície, realçando as variações macroestruturais causadas pelo processo de solidificação.
- Revelação da macroestrutura: Esse processo permite visualizar a região segregada e demais para análise visual por um especialista.

Ao final do processo de aquisição, foi obtido um conjunto de 180 imagens variando a resolução de 2.100 até 3.300 pixels, organizadas em três níveis distintos de acordo com a visibilidade da linha central resultante do processo de segregação. O nível 1 corresponde a amostras em que não há qualquer percepção da linha central; o nível 2 refere-se a amostras com percepção leve da linha central; e o nível 3 abrange imagens em que a linha central é

claramente visível. A Figura 2 apresenta exemplos de imagens macrográficas classificadas nos três níveis de segregação central.

Figura 2. Exemplos de imagens macrográficas dos níveis de segregação, a linha central não é vista no nível 1 e vista bem acentuada no nível 3.



(a) Nível 1

(b) Nível 2

(c) Nível 3

Há uma distribuição balanceada de amostras por classe, isto é, há um total de 60 imagens reais por classe. Para o posterior processamento das imagens pelo modelo proposto, houve a seguinte divisão dos dados: (i) 150 imagens para treinamento, distribuídas igualmente entre as três classes que representam os níveis de segregação (50 por classe); e (ii) 30 imagens reais para teste, com 10 imagens para cada classe.

3.2. Criação de Imagens com Modelos de Difusão

A utilização de redes neurais profundas para tarefas de visão computacional geralmente demanda grandes volumes de dados para atingir convergência satisfatória e métricas de desempenho elevadas. No entanto, nem sempre é viável a obtenção de novos dados reais. Com base nessa limitação, este trabalho propõe a geração de imagens sintéticas com o objetivo de auxiliar no processo de treinamento, promovendo tanto a aceleração da convergência quanto a melhoria das métricas de avaliação do modelo.

Para a geração das imagens sintéticas, optou-se pelo uso de modelos de difusão, os quais têm se destacado em relação a outras abordagens de geração de imagens, como as *Generative Adversarial Networks* (GANs) [Müller-Franzes et al. 2023]. Os modelos de difusão aprendem a gerar dados realistas ao reverter um processo de degradação de imagens induzido por ruído gaussiano progressivo [Ho et al. 2020]. Em termos gerais, esses modelos operam em duas etapas principais: (i) aplicação iterativa de ruído gaussiano a uma imagem original (processo de degradação); e (ii) aprendizado do processo inverso, no qual o modelo é treinado para reconstruir uma imagem coerente a partir de uma entrada ruidosa.

Neste trabalho, empregou-se o modelo *Stable Diffusion* [Chen and colleagues 2025] para a geração das imagens sintéticas. Ele realiza a reconstrução de imagens com base em descrições textuais, conhecidas como *prompts*. Para treinamento do modelo de difusão, utilizou-se o seguinte *prompt*:

a photo of a sks-segregation#ID steel plate

Já para a etapa de geração das imagens, utilizou-se o seguinte *prompt*:

a macrographic image of a sks-segregation#ID steel plate

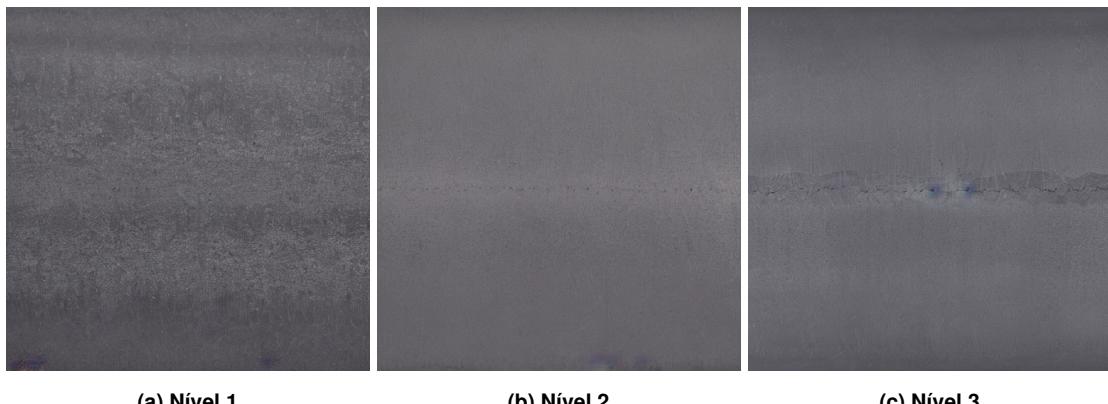
Em ambos os casos, os *prompts* são separados individualmente para cada nível da segregação, onde **#ID** é utilizado para este fim. Desta forma, um modelo de difusão é treinado para cada nível de segregação.

Vale destacar que o modelo gera imagens em alta resolução (512×512 pixels). Um aspecto relevante da arquitetura do *Stable Diffusion* é o uso do *Latent Diffusion Model* (LDM), no qual a difusão ocorre em um espaço latente comprimido ao invés do espaço de pixels. Para isso, é utilizado um *autoencoder* que comprime as imagens no espaço latente, onde ocorre o processo de difusão. Posteriormente, a imagem é decodificada para o espaço de pixels por meio de uma arquitetura U-Net convolucional. O *prompt* textual é codificado por meio do modelo CLIP, o que torna o processo computacionalmente mais eficiente.

Devido à escassez de imagens no conjunto original, foi empregada a técnica *DreamBooth* [Ruiz et al. 2023] para personalização do *prompt* e geração de imagens sintéticas específicas. O *DreamBooth* utiliza um pequeno conjunto de imagens representativas de um objeto ou conceito de interesse, rotuladas com um *prompt* da forma “*a [V] [super-category]*”, onde [V] é um *token* raro no vocabulário. O método realiza o ajuste fino da rede U-Net, e os autores relatam que três a cinco imagens já são suficientes para esse processo de personalização da geração texto-imagem.

Após o treinamento do modelo de difusão personalizado somente usando as imagens reservadas para treinamento, foram geradas 90 imagens sintéticas, distribuídas igualmente entre três classes (30 imagens por classe). A Figura 3 apresenta exemplos representativos das imagens geradas para cada classe.

Figura 3. Exemplos de imagens geradas pelo modelo de difusão para cada uma dos três níveis.



3.3. Treinamento da Rede Neural Convolucional

Antes mesmo de iniciar o treinamento, todas as imagens (sintéticas ou não) são redimensionadas para 224×224 e normalizadas entre 0.0 e 1.0 (divisão por 255). A escolha por este tamanho de entrada se deu pelo padrão presente na arquitetura usada neste trabalho: *EfficientNet B0* [Tan and Le 2019]. Isto permitiu o processo de *transfer learning* dos pesos pré-treinados da ImageNet.

A arquitetura final proposta utiliza a EfficientNet B0 como *backbone* e são adicionadas uma camada de GlobalAveragePooling2D, seguida por um *dropout* com taxa de 10% e duas camadas densas, uma com 128 neurônios e ativação ReLU e outra com três neurônios e ativação *softmax* para classificação, respectivamente.

O modelo foi inicialmente treinado utilizando as 150 imagens reais disponíveis e avaliado posteriormente sobre o conjunto de testes. Este resultado inicial é utilizado como *baseline* para comparação com os experimentos que incorporam dados sintéticos gerados por modelos de difusão.

Adicionalmente, com o intuito de estabelecer uma linha de base comparativa mais abrangente, também foram aplicadas técnicas tradicionais de *data augmentation*, como espelhamentos horizontais e verticais. Outros tipos de técnicas de aumento de dados não foram avaliadas pois iriam mudar a estrutura original da imagem (linha central na imagem que precisa ser detectada).

No experimento com dados sintéticos, as imagens geradas foram incorporadas ao conjunto de treinamento seguindo duas estratégias de avaliação: o modelo treinado com o *data augmentation* tradicional e sem. Em ambos os casos, foi realizado um refinamento (*fine-tuning*) utilizando todos os dados de treinamento.

3.4. Métricas de Avaliação das Amostras Classificadas

Sejam TP , TN , FP e FN os valores correspondentes aos verdadeiros positivos, verdadeiros negativos, falsos positivos e falsos negativos, respectivamente. Considerando que se trata de um problema de classificação multi-classe, adotam-se as seguintes definições para o cálculo dessas métricas em um esquema *one-vs-rest*: TP (Verdadeiro Positivo) é o número de amostras corretamente classificadas como pertencentes a uma determinada classe; TN (Verdadeiro Negativo) é o número de amostras corretamente classificadas como não pertencentes à classe em questão; FP (Falso Positivo) é o número de amostras incorretamente classificadas como pertencentes à classe em análise, quando na verdade pertencem a outra; e FN (Falso Negativo) corresponde ao número de amostras pertencentes à classe correta, mas classificadas como sendo de uma classe distinta.

As principais métricas utilizadas para avaliar o desempenho do classificador neste trabalho são:

- **Acurácia:** Trata-se de uma métrica que indica a fração de decisões corretas sobre o total de amostras avaliadas, definida por:

$$\text{Acurácia} = \frac{TP + TN}{TP + TN + FP + FN}. \quad (1)$$

- **Precisão:** Quantifica a confiabilidade das previsões positivas, definida por:

$$\text{Precisão} = \frac{TP}{TP + FP}. \quad (2)$$

- **Revocação:** A revocação mede a capacidade do modelo de encontrar todos os exemplos positivos, definida por:

$$\text{Revocação} = \frac{TP}{TP + FN}. \quad (3)$$

- **F1-Score:** Representa a média harmônica entre precisão e revocação, equilibrando ambos os extremos, definida por:

$$\text{F1-Score} = 2 \times \frac{\text{Precisão} \times \text{Revocação}}{\text{Precisão} + \text{Revocação}}. \quad (4)$$

Para este trabalho e considerando que é um problema multi-classe, utilizou-se a macro precisão, revocação e F1-score.

4. Experimentos e Resultados

Esta seção explica como foi conduzido o processo experimental deste trabalho. Mais especificamente, a Seção 4.1 apresenta a configuração experimental empregada; a Seção 4.2 apresenta os resultados obtidos; e a Seção 4.3 realiza uma discussão sobre os mesmos.

4.1. Setup de Experimentos

Para as execuções dos experimentos do trabalho, utilizou-se o Google Colab, além da linguagem Python. Para manipulação dos modelos foi utilizada o *framework* TensorFlow versão 2.18.0, já para o cálculo das métricas utilizou-se a biblioteca Scikit Learning versão 1.6.1.

Para a manipulação e criação das imagens sintéticas, utilizou-se PyTorch em conjunto com a biblioteca do HuggingFace nas versões 2.6.0 e 4.52.4, respectivamente, além do pacote Diffusers versão 0.33.1. Além disso, utilizou-se um lote de tamanho um para o treinamento, mas com a acumulação do gradiente a cada quatro. A taxa de aprendizado utilizada foi fixa de $5e^{-6}$ durante 1.000 iterações (uma imagem por iteração), e uma precisão mista de *float16* e o otimizador Adam.

O treinamento do modelo de classificação considerou uma taxa de aprendizado (*learning rate*) de 5×10^{-4} , utilizando o otimizador *Adam* e a função de perda *categorical crossentropy*, adequada para problemas de classificação multiclasse. Por questões de transparência e reproduzibilidade, o repositório de código utilizado neste trabalho está disponível publicamente.¹

4.2. Resultados

A Tabela 1 apresenta os resultados obtidos com o modelo proposto para a tarefa de classificação da segregação central em amostras de placas de aço. Para fins de comparação, o modelo foi avaliado em quatro configurações distintas: (i) sem a utilização de qualquer técnica de *data augmentation*; (ii) com a aplicação de estratégias tradicionais de *data augmentation*, incluindo reflexões horizontal e vertical; (iii) com a inclusão de dados sintéticos, mas sem *data augmentation*; e (iv) com dados sintéticos combinados às técnicas tradicionais de *data augmentation* previamente descritas.

No Cenário 1, utilizando-se exclusivamente 150 imagens reais, sem a aplicação de qualquer técnica de *data augmentation*, o modelo alcançou uma acurácia de 80,01%. A acurácia evidencia a limitação do modelo ao trabalhar apenas com dados reais, portanto, a importância de estratégias de aumento de dados para potencializar o desempenho base desse cenário é evidente durante a análise realizada.

¹Código disponível em <https://github.com/>.

Tabela 1. Comparação dos resultados entre os diferentes cenários. Acc = Acurácia; Pre = Precisão; Rev = Revocação; F1 = F1-score.

Cenário	Imagens Sintéticas	Data Augmentation	Acc (%)	Pre (%)	Rev (%)	F1
1	0	Não	80,01	79,56	80,00	0,7964
2	0	Sim	66,67	78,65	66,67	0,6604
3	90	Não	86,67	86,60	86,67	0,8648
4	90	Sim	93,33	94,44	93,33	0,9327

No Cenário 2, foram aplicadas técnicas tradicionais de aumento de dados como rotações aleatórias e reflexões horizontal e vertical ao mesmo conjunto de imagens reais. Contudo, observou-se uma leve redução na acurácia, que caiu para 66,67%. Este resultado sugere que, no contexto da classificação da segregação central, o uso de técnicas tradicionais de aumento de dados pode não ser suficiente para melhorar a capacidade de generalização do modelo. De fato, tais transformações podem introduzir ruído adicional nas imagens de treinamento, prejudicando a performance.

No Cenário 3, foram incorporadas 90 imagens sintéticas ao conjunto original de imagens reais, sem a aplicação de *data augmentation*. Esta configuração resultou em um ganho expressivo de desempenho, com acurácia de 86,67, evidenciando o potencial das imagens sintéticas como recurso complementar para aumentar a variabilidade do conjunto de dados e melhorar a eficácia do modelo.

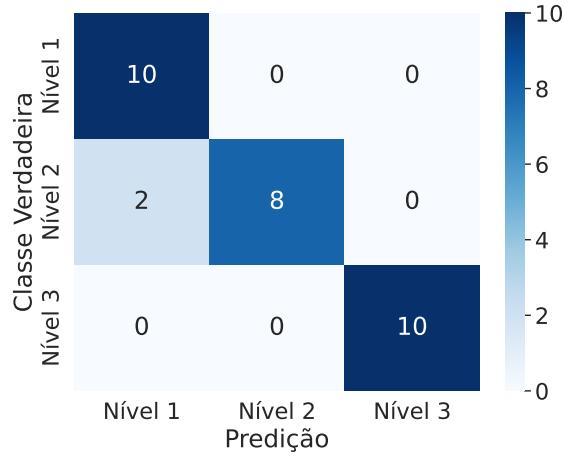
Por fim, o Cenário 4, que contempla a maior diversidade e volume de dados, composto por 150 imagens reais sem variações, 90 sintéticas, além da aplicação de *data augmentation* de reflexões horizontal e vertical e rotação de até 0,1 graus sobre as imagens reais, apresentou o melhor desempenho, alcançando acurácia de 93,33%. Este resultado reforça a relevância da diversidade e da quantidade de amostras para o bom desempenho de redes convolucionais. Além disso, confirma que abordagens que combinam dados reais, dados sintéticos e técnicas de enriquecimento são particularmente eficazes em cenários industriais caracterizados por restrições na obtenção de dados rotulados.

4.3. Discussão dos Resultados

A Figura 4 apresenta a matriz de confusão referente ao melhor modelo avaliado. Considerando o conjunto de teste composto por 30 amostras, o classificador obteve 28 acertos, correspondendo a uma acurácia global de 93,33%. Observa-se que todas as amostras pertencentes aos Níveis 1 e 3 foram corretamente classificadas. As únicas incorreções ocorreram no Nível 2, com duas instâncias rotuladas erroneamente como pertencentes ao Nível 1. Não foram observadas confusões envolvendo o Nível 3, o que sugere que esta classe apresenta características visuais suficientemente distintas. Em contrapartida, a sobreposição entre os padrões macrográficos dos Níveis 1 e 2 — especialmente nos casos em que a linha central característica do Nível 2 é sutil ou quase imperceptível — dificulta a separação precisa entre essas classes.

A Tabela 2 apresenta as métricas de precisão, revocação e F1-score para cada classe. O Nível 2 alcançou precisão de 100%, indicando ausência de falsos positivos, mas apresentou revocação de 80%, devido aos dois falsos negativos identificados na matriz de

Figura 4. Matriz de confusão obtida na etapa de teste.



confusão, resultando em um F1-score de 0,89. Para o Nível 1, a revocação foi de 100%, enquanto a precisão foi de 83%, uma vez que duas amostras do Nível 2 foram classificadas incorretamente como Nível 1. O Nível 3 apresentou desempenho perfeito em todas as métricas (100%), destacando-se como a classe mais facilmente identificável pelo modelo.

Tabela 2. Precisão, revocação e F1-score por classe do melhor modelo treinado.

Classe	Precisão (%)	Revocação (%)	F1-score
Nível 1	83.3	100.0	0.91
Nível 2	100.0	80.0	0.89
Nível 3	100.0	100.0	1.00

De forma geral, as médias macro obtidas foram de 94% para precisão, 93% para revocação e 0,93 para F1-score, refletindo um desempenho equilibrado entre confiabilidade e cobertura, mesmo em um cenário com distribuição de classes balanceada. É importante salientar que o terceiro nível (o com a linha central mais visível e que pode indicar possíveis problemas na fabricação) foi detectado em todos os casos, o que é extremamente importante para evitar a venda de produtos comprometidos.

5. Conclusão

Este trabalho teve como objetivo propor e avaliar uma abordagem de classificação automática da segregação central em placas de aço, utilizando redes neurais convolucionais aliadas a estratégias de aumento de dados baseadas em imagens reais e sintéticas. A principal contribuição foi demonstrar que a combinação de dados sintéticos gerados por modelos de difusão com técnicas tradicionais de *data augmentation* pode ampliar a diversidade e a representatividade do conjunto de treinamento, resultando em melhoria da acurácia de 80,01% para 93,33%. O método proposto apresenta baixo custo computacional, podendo ser facilmente incorporado a processos de inspeção visual automatizada em ambientes industriais.

Como perspectivas futuras, sugere-se investigar outros métodos de geração de dados sintéticos, como *Textual Inversion*, e testar estratégias de *data augmentation* mais

avançadas, visando ampliar ainda mais a capacidade de generalização do sistema. Também se recomenda avaliar outros tipos de defeitos analisados via imagens macrográficas, permitindo a análise automática e imediata das amostras, contribuindo para maior agilidade e confiabilidade no controle de qualidade siderúrgico.

Agradecimentos

Os autores gostariam de agradecer à Fundação de Amparo à Pesquisa do Estado de Minas Gerais (FAPEMIG, subsídio APQ-01768-24), à Coordenação de Aperfeiçoamento de Pessoal de Nível Superior (CAPES), ao Conselho Nacional de Desenvolvimento Científico e Tecnológico (CNPq), à Universidade Federal de Ouro Preto (UFOP/PROPPI) e à Gerdau S/A por apoiar o desenvolvimento do presente estudo.

Referências

- Beckermann, C. (2002). Modelling of macrosegregation: Applications and future needs. *International Materials Reviews*, 47(5):243–261.
- Chen, Z. and colleagues (2025). Stable diffusion models are secretly good at visual in-context learning. In *International Conference on Learning Representations (ICLR)*. <https://openreview.net/forum?id=fKrFTGnoXY>.
- Costa, A. B. S., Nishiura, J. Y., Lopes, P. V., Verri, F. A. N., and Skoogh, A. (2024). Artificial data generation for smart manufacturing systems: Discrete event simulation, traceability, and process mining. In *Encontro Nacional de Inteligência Artificial e Computacional (ENIAC)*, pages 719–730. SBC.
- Fu, G., Zhang, Z., Le, W., Li, J., Zhu, Q., Niu, F., Chen, H., Sun, F., and Shen, Y. (2023). A multi-scale pooling convolutional neural network for accurate steel surface defects classification. *Frontiers in Neurorobotics*, 17:1096083.
- Gao, Y., Gao, L., Li, X., and Yan, X. (2019). A semi-supervised convolutional neural network-based method for steel surface defect recognition. *Robotics and Computer-Integrated Manufacturing*, 61:101825.
- Ho, J., Jain, A., and Abbeel, P. (2020). Denoising diffusion probabilistic models. In *Advances in Neural Information Processing Systems*, volume 33, pages 6840–6851.
- Hu, Y., Tang, J., Xu, Y., Xu, R., and Huang, B. (2024). EL-DenseNet: A novel method for identifying the flame state of converter steelmaking based on dense convolutional neural networks. *Signal, Image and Video Processing*, 18:3445–3457.
- Ibrahim, A. A. M. (2024). Transfer learning-based approach using new convolutional neural network classifier for steel surface defects classification. *Scientific African*, 23:e02066.
- Li, H., Liu, Y., Liu, C., Pang, H., and Xu, K. (2025). A few-shot steel surface defect generation method based on diffusion models. *Sensors*, 25(10):3038. Available at: <https://doi.org/10.3390/s25103038>.
- Lima, F. R. and Gomes, R. (2020). Conceitos e tecnologias da indústria 4.0: uma análise bibliométrica. *Revista Brasileira de Inovação*, 19:e0200023.
- Ludwig, A., Wu, M., and Kharicha, A. (2015). On macrosegregation. *Metallurgical and Materials Transactions A*, 46:4854–4867.

- Moore, J. (1984). Review of axial segregation in continuous cast steel. *Continuous Casting, Iron and Steel Society*, 3:11–20.
- Müller-Franzes, G., Niehues, J. M., Khader, F., Arasteh, S. T., Haarburger, C., Kuhl, C., Wang, T., Han, T., Nolte, T., Nebelung, S., et al. (2023). A multimodal comparison of latent denoising diffusion probabilistic models and generative adversarial networks for medical image synthesis. *Scientific Reports*, 13(1):12098.
- Nieto, P. J. G., García-Gonzalo, E., Álvarez Antón, J. C., Suárez, V. M. G., Bayón, R. M., and Martín, F. M. (2018). A comparison of several machine learning techniques for the centerline segregation prediction in continuous cast steel slabs and evaluation of its performance. *Journal of Computational and Applied Mathematics*, 330:877–895.
- Ottoni, A. L. C., de Amorim, R. M., Novo, M. S., and Costa, D. B. (2023). Tuning of data augmentation hyperparameters in deep learning to building construction image classification with small datasets. *International Journal of Machine Learning and Cybernetics*, 14(1):171–186.
- Queiroz, L. C. L. (2015). Comparison of edge detection techniques applied in the identification of centerline segregation on steel slabs. *American Journal of Applied Sciences*, 12(8):567–571.
- Ruiz, N., Li, Y., Jampani, V., Pritch, Y., Rubinstein, M., and Aberman, K. (2022). Dreambooth: Fine tuning text-to-image diffusion models for subject-driven generation.
- Ruiz, N., Li, Y., Jampani, V., Pritch, Y., Rubinstein, M., and Aberman, K. (2023). Dreambooth: Fine tuning text-to-image diffusion models for subject-driven generation. In *Proceedings of the IEEE/CVF conference on computer vision and pattern recognition*, pages 22500–22510.
- Shorten, C. and Khoshgoftaar, T. M. (2019). A survey on image data augmentation for deep learning. *Journal of Big Data*, 6(1):60.
- Tai, Y., Yang, K., Peng, T., Huang, Z., and Zhang, Z. (2021). Defect image sample generation with diffusion prior for steel surface defect recognition. *Journal of LaTeX Class Files*, 14(8):1–13. Available at: arXiv:2405.01872 [cs.CV].
- Tan, M. and Le, Q. (2019). Efficientnet: Rethinking model scaling for convolutional neural networks. In Chaudhuri, K. and Salakhutdinov, R., editors, *Proceedings of the 36th International Conference on Machine Learning*, volume 97 of *Proceedings of Machine Learning Research*, pages 6105–6114. PMLR.
- Zhang, C., Gao, W., Ma, C., Zhang, R., Yin, H., and Du, Y. (2024). Diffusion-based text-to-image generative model for predicting materials microstructure evolutions. *SSRN Electronic Journal*. Preprint, available at SSRN: <https://ssrn.com/abstract=5248820>.
- Zhang, R. and Yang, J. (2023). State of the art in applications of machine learning in steelmaking process modeling. *International Journal of Minerals, Metallurgy and Materials*, 30(11):2055–2067.
- Zou, L., Zhang, J., Liu, Q., Zeng, F., Chen, J., and Guan, M. (2019). Prediction of central carbon segregation in continuous casting billet using a regularized extreme learning machine model. *Metals*, 9(12):1312.