

FRANCISCO LEONARDO AURELIANO FERREIRA

**MODELO DE REDES NEURAIIS SIAMESAS  
PARA O RECONHECIMENTO DE  
CARACTERES MANUSCRITOS**

Mossoró-RN

2024



UNIVERSIDADE FEDERAL RURAL DO SEMI-ÁRIDO  
UNIVERSIDADE DO ESTADO DO RIO GRANDE DO NORTE  
PROGRAMA DE PÓS-GRADUAÇÃO EM CIÊNCIA DA  
COMPUTAÇÃO



MESTRANDO

FRANCISCO LEONARDO AURELIANO FERREIRA

ORIENTADOR

Dr. LENARDO CHAVES E SILVA

COORIENTADOR

Dr. ÁLVARO ALVARES DE CARVALHO CÉSAR SOBRINHO

MODELO DE REDES NEURAIAS SIAMESAS PARA O  
RECONHECIMENTO DE CARACTERES  
MANUSCRITOS

Mossoró-RN

2024

**FRANCISCO LEONARDO AURELIANO FERREIRA**

**MODELO DE REDES NEURAIS SIAMESAS PARA O  
RECONHECIMENTO DE CARACTERES  
MANUSCRITOS**

Dissertação apresentada ao Programa de Pós-Graduação em Ciência da Computação - associação ampla entre a Universidade do Estado do Rio Grande do Norte e a Universidade Federal Rural do Semi-Árido, para a obtenção do título de Mestre em Ciência da Computação.

Orientador: Prof<sup>o</sup> Dr. LENARDO CHAVES E SILVA

Coorientador: Prof<sup>o</sup> Dr. ÁLVARO ALVARES DE CARVALHO CÉSAR SOBRINHO

**Mossoró-RN**

**2024**

**FRANCISCO LEONARDO AURELIANO FERREIRA**

**MODELO DE REDES NEURAIS SIAMESAS PARA O  
RECONHECIMENTO DE CARACTERES  
MANUSCRITOS**

Dissertação apresentada ao Programa de Pós-Graduação em Ciência da Computação para a obtenção do título de Mestre em Ciência da Computação.

APROVADA EM: \_\_\_\_ / \_\_\_\_ / \_\_\_\_.

BANCA EXAMINADORA

---

**Prof<sup>o</sup> Dr. LENARDO CHAVES E  
SILVA**  
Orientador

---

**Avaliador Externo, Dsc.**  
Avaliador Externa

---

**Avaliador do Programa, Dsc.**  
Avaliador do Programa

---

**Professor**  
Convidado 3

Dedico esse trabalho aos meus pais, Maria e José,  
a minha vó Perpetua e a todos os meus amigos.

# Agradecimentos

Primeiramente, agradeço a DEUS por me conceder força, sabedoria e resiliência ao longo dessa jornada. À minha família, minha eterna gratidão por me ensinar que, através da educação, seria possível transformar minha vida e a vida deles. Obrigado pelo apoio incondicional, motivação contínua e por sempre acreditarem em mim.

Aos meus amigos, que estiveram ao meu lado durante os momentos desafiadores e comemoraram comigo cada conquista, agradeço pela amizade, colaboração e pelas valiosas trocas de ideias.

Agradeço imensamente ao meu orientador, Dr. LENARDO CHAVES, por todo o conhecimento transmitido, paciência e orientação ao longo deste trabalho. Sua expertise foi essencial para a realização deste estudo.

Ao meu coorientador, Dr. ÁLVARO SOBRINHO, agradeço profundamente pela ajuda inestimável, pelas discussões esclarecedoras e pelo apoio técnico que tanto contribuíram para o desenvolvimento desta pesquisa.

Finalmente, expresso minha gratidão a todos que, de alguma forma, contribuíram para a realização deste trabalho. A todos vocês, meu sincero muito obrigado!

# Resumo

A exclusão digital continua a ser um desafio significativo na educação, especialmente em regiões com acesso limitado a recursos tecnológicos. Nesse cenário, a avaliação automática de redações manuscritas por meio de *Inteligência Artificial (IA)* surge como uma solução promissora para apoiar educadores e alunos. Este trabalho desenvolveu um modelo de *Redes Neurais Convolucionais Siamesas (SCNNs)* utilizando aprendizado de poucos exemplos (*few-shot learning*) para o reconhecimento de caracteres manuscritos em 440 redações de alunos do ensino fundamental público no Brasil. Para lidar com o desequilíbrio nas quantidades de imagens das diferentes classes de caracteres, foram aplicadas técnicas de aumento de dados. Métodos de segmentação de caracteres e rotulagem manual foram empregados para criar um conjunto de dados representativo e diversificado. O modelo proposto, utilizando a função de perda *Entropy Loss*, alcançou uma precisão entre 80% e 85% no reconhecimento de caracteres. Experimentos com diferentes bases de dados, incluindo *EMNIST*, *NIST19* e *ImageNet*, demonstraram a capacidade do modelo de se adaptar a novos exemplos com eficácia. Este estudo contribui para a *IA* na educação ao oferecer uma abordagem que melhora a legibilidade dos caracteres manuscritos, auxiliando na identificação e correção de problemas de escrita. A implementação dessa tecnologia em contextos educacionais pode ser particularmente benéfica em países de baixa e média renda, onde as escolas enfrentam desafios socioeconômicos e tecnológicos. A melhoria da legibilidade não só facilita a avaliação automática de redações, mas também contribui para o desenvolvimento das habilidades de escrita dos alunos, promovendo uma aprendizagem mais eficaz e inclusiva.

**Palavras-chave:** Educação, Reconhecimento de Escrita, Inteligência Artificial.

# Abstract

Digital exclusion remains a significant challenge in education, especially in regions with limited access to technological resources. In this context, the automatic evaluation of handwritten essays through *Artificial Intelligence (AI)* emerges as a promising solution to support educators and students. This study developed a model based on *Siamese Convolutional Neural Networks (SCNNs)* using *few-shot learning* for the recognition of handwritten characters in 440 essays from public elementary school students in Brazil. To address the imbalance in the number of images across different character classes, data augmentation techniques were applied. Character segmentation methods and manual labeling were employed to create a representative and diverse dataset. The proposed model, using the *Entropy Loss* function, achieved an accuracy between 80% and 85% in character recognition. Experiments with different datasets, including *EMNIST*, *NIST19*, and *ImageNet*, demonstrated the model's ability to effectively adapt to new examples. This study contributes to *AI* in education by offering an approach that enhances the legibility of handwritten characters, assisting in the identification and correction of writing issues. Implementing this technology in educational contexts can be particularly beneficial in low- and middle-income countries, where schools face socio-economic and technological challenges. Improving legibility not only facilitates the automatic evaluation of essays but also contributes to the development of students' writing skills, promoting more effective and inclusive learning.

**Keywords:** Education, Handwriting Recognition, Artificial Intelligence.



# Lista de ilustrações

Figura 1 – Representação da metodologia de pesquisa pela perspectiva do <i>Design Science</i> . . . . .	17
Figura 2 – Etapas para o reconhecimento óptico de caracteres. . . . .	20
Figura 3 – Visão geral do processo de reconhecimento de escrita. Adaptado de Choudhary et al. (2017). . . . .	21
Figura 4 – Arquitetura proposta por (MCLAUGHLIN; RINCON; MILLER, 2016), que consiste em uma rede neural siamesa composta por CNN e RNN. . . . .	24
Figura 5 – Arquitetura simplificada de uma rede siamesa (Fonte: (CHAVES <i>et al.</i> , 2021)). . . . .	26
Figura 6 – Passos metodológicos para implementar e validar o modelo. . . . .	38
Figura 7 – Pré-processamento manual e automático. . . . .	40
Figura 8 – Exemplos de imagens segmentadas. . . . .	42
Figura 9 – Conjunto de suporte. . . . .	43
Figura 10 – Amostras de letras e dígitos do conjunto de dados EMNIST. . . . .	45
Figura 11 – Amostras do conjunto de dados ImageNet. . . . .	46
Figura 12 – Acurácia em relação ao número de épocas para diferentes bases e funções de perda . . . . .	63
Figura 13 – Precisão para EMNIST com Contrastive Loss e Entropy Loss . . . . .	65
Figura 14 – Recall para EMNIST com Contrastive Loss e Entropy Loss . . . . .	67
Figura 15 – Comparação de Acurácia entre as Funções de Perda <i>Entropy Loss</i> e <i>Contrastive Loss</i> . . . . .	70

# Lista de tabelas

Tabela 1 – Modelo Siamesa em Diferentes Áreas e Objetivos . . . . .	33
Tabela 2 – Conjuntos de Dados Utilizados nos Experimentos . . . . .	44
Tabela 3 – Configuração do Sistema . . . . .	50
Tabela 4 – Hiperparâmetros e Técnicas Utilizadas no Treinamento da SCNN . . . . .	56
Tabela 5 – Comparação entre o modelo original e o modelo proposto . . . . .	60
Tabela 6 – Acurácia do modelo proposto no conjunto de dados EMNIST. . . . .	61
Tabela 7 – Acurácia do modelo proposto no conjunto de dados NIST19. . . . .	62
Tabela 8 – Acurácia do modelo proposto no conjunto de dados ImageNET. . . . .	63
Tabela 9 – Comparação das métricas de desempenho para os conjuntos de dados EMNIST, NIST19 e ImageNET . . . . .	64
Tabela 10 – Ranking de Acurácias Obtidas nos Experimentos com a Rede Neural Siamesa . . . . .	69
Tabela 11 – Resultados do Teste de Hipótese . . . . .	70
Tabela 12 – Comparação dos Trabalhos Relacionados . . . . .	72
Tabela 13 – Aplicabilidade do Modelo Siamesa em Diferentes Áreas . . . . .	74

# Lista de abreviaturas e siglas

<i>AI</i>	<i>Artificial Intelligence</i>
<i>BDE</i>	Banco de Dados Educacionais
<i>CE</i>	Critério de Exclusão
<i>CI</i>	Critério de Inclusão
<i>CNN</i>	<i>Convolutional Neural Network</i>
<i>DS</i>	<i>Design Science</i>
<i>HMM</i>	<i>Hidden Markov Model</i>
<i>IA</i>	Inteligência Artificial
<i>kNN</i>	<i>k-Nearest Neighbors</i>
<i>LSTM</i>	<i>Long Short Term Memory</i>
<i>ML</i>	<i>Machine Learning</i>
<i>MTCNN</i>	<i>Multi-task Cascaded Convolutional Networks</i>
<i>OCR</i>	<i>Optical Character Recognition</i>
<i>PDI</i>	Processamento Digital de Imagem
<i>PLN</i>	Processamento de Linguagem Natural
<i>Q1</i>	Questão 1
<i>Q2</i>	Questão 2
<i>Q3</i>	Questão 3
<i>QPP</i>	Questão Principal de Pesquisa
<i>Qp</i>	Questão de Pesquisa
<i>RNCS</i>	Rede Neural Convolutacional Siamesa
<i>RSL</i>	Revisão Sistemática da Literatura
<i>TF-IDF</i>	<i>Term Frequency-Inverse Document Frequency</i>

# Sumário

<b>1</b>	<b>INTRODUÇÃO</b>	<b>12</b>
1.1	Motivação	14
1.2	Questões de Pesquisa	15
1.3	Objetivos da Pesquisa	15
1.4	Metodologia	16
1.5	Organização do Documento	18
<b>2</b>	<b>FUNDAMENTAÇÃO TEÓRICA</b>	<b>19</b>
2.1	Reconhecimento de Escrita	19
2.2	Redes Neurais Convolucionais	24
2.2.1	Redes Neurais Convolucionais Siamesas	25
2.3	Few-Shot Learning	27
2.3.1	Vantagens e Desafios do Few-Shot Learning	28
2.3.2	Aplicações do FSL	28
<b>3</b>	<b>TRABALHOS RELACIONADOS</b>	<b>30</b>
<b>4</b>	<b>MATERIAIS E MÉTODOS</b>	<b>38</b>
4.1	Problema Principal	38
4.2	Coleta de Dados e Pré-processamento	38
4.2.1	Segmentação e construção de conjunto de suporte	41
4.3	Aquisição da base de dados	44
4.3.1	Conjunto de Dados EMNIST	44
4.3.2	Conjunto de Dados NIST Special Database 19	45
4.3.3	Conjunto de Dados IMAGENET	45
4.4	Implementação do Modelo.	46
4.4.1	Etapas de Implementação	47
4.4.1.1	Estrutura do Few-Shot Learning	48
4.4.1.2	Processo de Few-Shot Learning	48
4.4.1.3	Cenário Few-Shot Learning no modelo proposto	49
4.4.2	Ambiente de trabalho	49
4.5	Rede Neural Proposta	50
4.5.1	Rede Neural Convolucional Siamesa	50
4.5.2	Arquitetura da Rede Neural Siamesa	51
4.5.3	Parâmetros e Hiperparâmetros	52

<b>5</b>	<b>EXPERIMENTOS E RESULTADOS</b>	<b>53</b>
<b>5.1</b>	<b>Configuração Experimental</b>	<b>54</b>
<b>5.2</b>	<b>Resultados</b>	<b>60</b>
5.2.1	Experimentos com EMNIST	61
5.2.2	Experimentos com NIST19	62
5.2.3	Experimentos com ImageNET	62
<b>6</b>	<b>DISCUSSÃO</b>	<b>68</b>
<b>6.1</b>	<b>Discussão dos resultados obtidos</b>	<b>68</b>
6.1.1	Significância Estatística dos Resultados	69
6.1.2	Implicações dos Resultados	71
6.1.3	Diferenciação do Trabalho Desenvolvido	71
6.1.4	Limitações e Trabalhos Futuros	73
6.1.5	Aplicabilidade do Modelo Proposto	73
<b>7</b>	<b>CONCLUSÕES</b>	<b>76</b>
	<b>REFERÊNCIAS</b>	<b>78</b>

# 1 Introdução

Nos dias atuais, o uso de ferramentas capazes de interpretar a linguagem natural e realizar reconhecimento de manuscritos e avaliação automática tem se tornado cada vez mais comum (SILVA *et al.*, 2024). Esses avanços são possíveis graças ao desenvolvimento na área da Inteligência Computacional, incluindo o Processamento da Linguagem Natural (PLN) e a Visão Computacional. A desigualdade educacional tem sido um problema crítico em todo o mundo, especialmente no Sul Global, onde a pandemia de COVID-19 agravou ainda mais os resultados de aprendizagem para 1,6 bilhão de estudantes globalmente (HELEN, 2022). Competências essenciais do século XXI, como a comunicação oral e escrita, foram negativamente impactadas pela desigualdade educacional. Os países do Sul Global enfrentam vários desafios que contribuem para essa desigualdade, incluindo a falta de infraestrutura, treinamento inadequado, dificuldades no apoio a áreas remotas, baixa qualidade do material e escassez de recursos (UNESCO, 2020).

Conceitos emergentes, como a Inteligência Artificial (IA) na Educação Desplugada, têm surgido para abordar essas questões. A abordagem "AIED Unplugged" foi proposta para mitigar a exclusão digital, especialmente em regiões desfavorecidas. Esta abordagem tem como foco a criação de tecnologias educacionais baseadas em IA que não requerem infraestrutura avançada, acesso estável à internet ou habilidades digitais dos usuários, promovendo assim a inclusão educacional em larga escala (ISOTANI *et al.*, 2023). No Brasil, o conceito de IA na Educação Desplugada foi aplicado com sucesso em um estudo de caso que melhorou as habilidades de escrita em escolas primárias no estado do Amazonas (PORTELA; COLABORADORES, 2023). Nessa abordagem, a caligrafia dos alunos é capturada dentro da sala de aula usando um smartphone de baixo custo. As imagens coletadas são transmitidas quando há conectividade com a internet, permitindo a avaliação automática offline das produções de texto. Esta abordagem é especialmente relevante no Sul Global, onde muitos alunos em escolas públicas não têm acesso a computadores para digitação devido às suas vulnerabilidades socioeconômicas. Diferentemente de muitas regiões do mundo onde o acesso a computadores é mais comum e a digitação é amplamente utilizada, no Sul Global, a escrita manual ainda é predominante nas escolas, devido às limitações econômicas e tecnológicas (GAŠEVIĆ, 2018).

O uso da IA na educação tem abordado diversos problemas relacionados ao reconhecimento de escrita e análise textual (FERREIRA-MELLO *et al.*, 2019), como pontuação de redação (KUMAR; BOULANGER, 2021), análise de coesão e coerência (VAJJALA, 2018), avaliação da estrutura retórica de textos (MELLO *et al.*, 2021) e avaliação de respostas curtas (ORMEROD *et al.*, 2022). Por exemplo, (VASHIST; PANDEY; TRIPATHI, 2020) compararam técnicas de reconhecimento de escrita com foco em aprendizado de máquina.

Além disso, (REMAIDA *et al.*, 2020) revisaram a literatura sobre reconhecimento de escrita usando redes neurais artificiais.

As técnicas de IA podem apoiar o reconhecimento de caligrafia, possibilitando aplicações como pontuação de redações, análise de coesão e coerência, avaliação da estrutura retórica em textos e avaliação de respostas curtas. O reconhecimento de caligrafia é desafiador devido ao estilo de escrita único de cada aluno e à legibilidade do texto. As imagens de redações capturam a caligrafia real dos alunos, permitindo uma avaliação mais precisa e contextualizada de suas habilidades de escrita. Isso é particularmente relevante em países com acesso limitado a recursos tecnológicos, onde a avaliação digital de textos manuscritos pode ser uma solução prática e eficiente. A abordagem AIED desplugada promove o uso de IA em ambientes educacionais sem a necessidade de infraestrutura digital avançada, tornando-a acessível a escolas públicas brasileiras. No entanto, a disponibilidade limitada de conjuntos de dados com imagens de redações de alunos brasileiros é um desafio, pois requer a consideração de características linguísticas e estilísticas específicas do Brasil.

O reconhecimento de caracteres manuscritos é um desafio significativo devido às variações no estilo de escrita de cada indivíduo e às condições variadas em que as escritas são produzidas. A legibilidade dos caracteres é crucial para garantir a precisão na interpretação e análise dos textos manuscritos. Redes Neurais Convolucionais Siamesas (SCNNs) têm se mostrado eficazes nesse contexto por várias razões. As SCNNs são particularmente adequadas para cenários onde há uma quantidade limitada de dados rotulados, como é o caso do contexto educacional, onde pode ser difícil coletar grandes volumes de dados de escrita de estudantes. Essa abordagem permite que o modelo aprenda a identificar caracteres com base em um número muito pequeno de exemplos. As SCNNs são projetadas para identificar similaridades entre pares de entradas, o que significa que o modelo pode aprender a reconhecer caracteres mesmo que eles apresentem variações significativas em termos de estilo, tamanho e orientação. Ao treinar o modelo com dados que incluem uma variedade de estilos de escrita, as SCNNs podem ajudar a melhorar a legibilidade geral dos textos manuscritos. Isso é especialmente útil em ambientes educacionais, onde o feedback imediato sobre a clareza e a forma da escrita pode ser utilizado para ajudar os alunos a melhorar suas habilidades de escrita.

A abordagem proposta neste estudo visa não apenas melhorar a precisão do reconhecimento de caracteres manuscritos, mas também abordar os desafios específicos enfrentados no contexto educacional brasileiro. Ao focar na legibilidade e na robustez do modelo em relação a variações na escrita, espera-se proporcionar uma ferramenta útil para educadores e estudantes, contribuindo para a melhoria da qualidade da educação e para a redução das desigualdades educacionais.

Frente a esse cenário, este trabalho apresenta uma Revisão da Literatura (RL), investigando abordagens utilizadas no reconhecimento de escrita, tipos de técnicas com-

putacionais empregadas, objetivos educacionais dessas ferramentas, soluções propostas e conjuntos de dados disponíveis, entre outros aspectos. Além disso, realiza-se um estudo sobre a utilização de SCNNs e aprendizagem de poucos exemplos para reconhecimento de caracteres a partir de imagens de redações manuscritas de alunos do ensino fundamental público no Brasil.

## 1.1 Motivação

A principal motivação deste trabalho é enfrentar a dificuldade em fornecer métodos automatizados para a correção e melhoria da escrita de estudantes de escolas públicas. No contexto educacional, especialmente em países em desenvolvimento, muitos alunos não têm acesso a ferramentas digitais que possam auxiliar na aprendizagem e na melhoria da caligrafia. Isso cria um desafio significativo para educadores e gestores que buscam formas eficazes de avaliar e melhorar a escrita dos alunos de maneira eficiente e escalável.

Este trabalho busca desenvolver uma ferramenta baseada em Redes Neurais Convolucionais Siamesas (SCNNs) e aprendizado de poucos exemplos (few-shot learning) para o reconhecimento de caracteres manuscritos. A proposta é criar um modelo capaz de funcionar eficazmente mesmo com uma quantidade limitada de exemplos de treinamento, o que é crucial em ambientes onde a coleta de grandes volumes de dados rotulados é desafiadora.

Utilizando técnicas avançadas de aprendizado de máquina, este estudo visa proporcionar uma solução prática e acessível que possa ser aplicada diretamente no contexto educacional. A ferramenta desenvolvida tem o potencial de:

- **Auxiliar na correção automática de redações:** Oferecendo feedback imediato sobre a legibilidade e forma dos caracteres escritos pelos alunos.
- **Apoiar educadores:** Fornecendo uma maneira eficiente de avaliar a escrita dos alunos e identificar áreas que necessitam de melhoria.
- **Promover a inclusão digital:** Garantindo que alunos de escolas públicas tenham acesso a tecnologias avançadas de aprendizado de máquina, independentemente de suas condições socioeconômicas.

A motivação para este trabalho surge da necessidade de inovar na maneira como avaliamos e melhoramos a escrita dos alunos, utilizando ferramentas tecnológicas que possam ser amplamente acessíveis e eficazes.



## 1.2 Questões de Pesquisa

O presente trabalho segue o paradigma de Design Science (DS) (PIMENTEL; FILIPPO; SANTOS, 2020), com o objetivo de gerar novos conhecimentos técnicos e científicos, no qual as questões de pesquisa produzidas conduzem a metodologia de pesquisa e são respondidas no decorrer do desenvolvimento da pesquisa. Diante desse contexto, o estudo tem como foco a utilização de Redes Neurais Convolucionais Siamesas (SCNNs) e aprendizagem com poucas amostras para reconhecimento de caracteres a partir de imagens de ensaios manuscritos de alunos do ensino fundamental público no Brasil. De acordo com essa perspectiva, a Questão de Pesquisa Principal (QPP) a ser respondida é:

**QPP: Quais são as abordagens utilizadas e qual o modelo proposto com melhor desempenho no reconhecimento de escrita?**

**QP1** Até que ponto a utilização de SCNN com poucas amostras é capaz de reconhecer caracteres a partir de imagens de redações manuscritas?

**QP2** Como essa técnica se compara com outras técnicas existentes no reconhecimento de escrita manuscrita?

**QP3** Como a função de perda Entropy Loss se compara com a função de perda Contrastive Loss utilizando a SCNN em tarefas de *few-shot learning*?

## 1.3 Objetivos da Pesquisa

O objetivo geral deste trabalho é realizar um estudo e experimentação sobre a utilização de (SCNNs) e aprendizagem com poucas amostras (*few-shot learning*) para o reconhecimento de caracteres a partir de imagens de redações manuscritas de alunos do ensino fundamental público no Brasil. Para alcançar este objetivo, vários objetivos específicos precisam ser atingidos, incluindo:

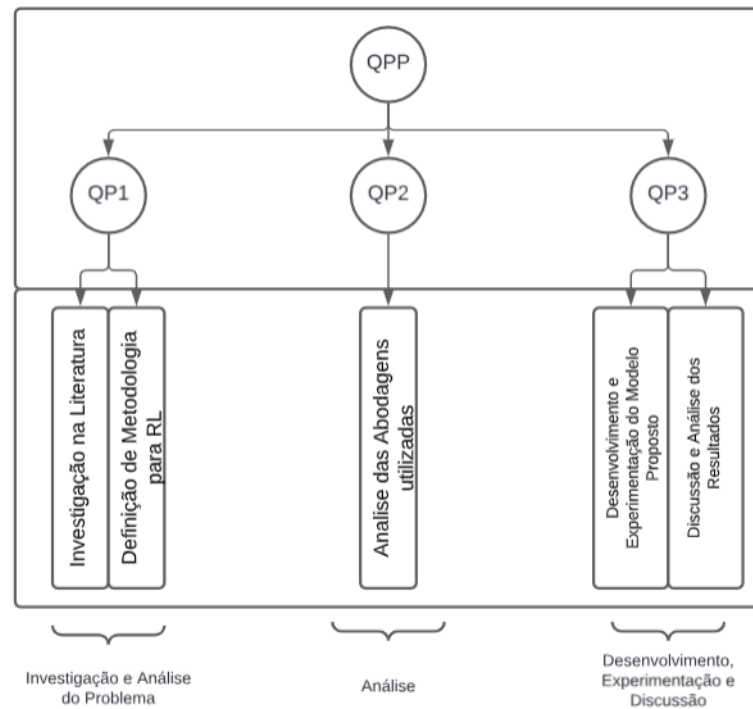
1. **Investigar o Estado da Arte:** realizar uma revisão da literatura (RL) para identificar as técnicas, modelos, algoritmos e ferramentas utilizadas no reconhecimento de escrita em contextos educacionais.
2. **Analisar Técnicas de Reconhecimento de Caracteres:** realizar uma análise detalhada das principais técnicas utilizadas no reconhecimento de caracteres manuscritos, identificando seus pontos fortes e limitações.
3. **Implementar e Testar Modelos:** desenvolver e implementar modelos de redes neurais siamesas, ajustando hiperparâmetros e utilizando diferentes funções de perda para avaliar o desempenho no reconhecimento de caracteres manuscritos.

4. **Aplicar Aprendizagem com Poucas Amostras:** utilizar técnicas de *few-shot learning* para treinar o modelo com bases de dados contendo poucas amostras, para que o modelo possa generalizar bem mesmo com um número limitado de exemplos de treino.
5. **Avaliar Desempenho:** avaliar o desempenho dos modelos propostos utilizando métricas como precisão, *recall*, *F1-score* e acurácia, e comparar com outras abordagens existentes.
6. **Discutir Resultados:** apresentar uma discussão detalhada sobre os resultados obtidos, incluindo uma análise crítica dos desafios e benefícios das abordagens utilizadas, especialmente no contexto de bases de dados com poucas amostras, e sugestões para trabalhos futuros.

## 1.4 Metodologia

Esta pesquisa caracteriza-se como teórico-aplicada, pois são estudados conceitos, técnicas e métodos que envolvem o estudo e comparação de abordagens utilizadas no processo de reconhecimento de escrita, assim como, por ser amplamente utilizado na pesquisa de sistemas de informação para resolver problemas complexos. O DS foi escolhido, por ter também emergido como um método apropriado para diversas investigações no campo da educação, como demonstrado por trabalhos que utilizam DS para analisar dados educacionais e melhorar processos de ensino e aprendizagem (KLAŠNJA-MILIĆEVIĆ, 2017). Por exemplo, (TSIAKMAKI, 2021) utilizaram métodos de DS para prever o desempenho acadêmico dos alunos, enquanto (PINARGOTE-ORTEGA *et al.*, 2019) aplicaram técnicas de DS para avaliação entre pares em ambientes educacionais. Foram formuladas questões que conduzem a metodologia de pesquisa, no qual as mesmas são respondidas ao longo da execução da pesquisa.

Diante desta perspectiva, o conjunto problema é formado por questões secundárias QP1, QP2 e QP3. Já as soluções representam as principais contribuições do trabalho, alcançadas por meio da investigação do problema no estado da arte, análise, experimentação e por fim, discussão do experimento e resultado. Como mostra a Figura 1 a QPP dá origem às questões secundárias, as quais fazem parte de três etapas: investigação e análise do problema, análise e desenvolvimento, experimentação e discussão.

Figura 1 – Representação da metodologia de pesquisa pela perspectiva do *Design Science*.

Fonte: Autoria própria.

De acordo com (CIRIBELLI, 2003), a metodologia científica pode ser entendida como um conjunto de etapas e instrumentos pelo qual o pesquisador, direciona seu projeto de trabalho com critérios de caráter científico para alcançar dados que dão ao conjunto problema, direcionamento ao conjunto solução. Neste trabalho foram utilizados três etapas, descritas a seguir.

**Etapa 1: REVISÃO DA LITERATURA:** Essa etapa pode ser entendida como a junção de ideias de diferentes autores sobre determinado tema, conseguidas através de leituras, de pesquisas realizadas pelo pesquisador. É realizado a pesquisa nas fontes de buscas para a elaboração do referencial teórico do trabalho. Segundo (DRESCH; LACERDA; JUNIOR, 2015), RL é uma etapa fundamental da condução de pesquisas científicas, especialmente de pesquisas realizadas sob o paradigma da DS. A RL segue algumas etapas, que o pesquisador precisa entender e seguir para que o trabalho de revisão seja bem feito, tendo em vista a minimização dos problemas que podem atrapalhar, ou mesmo deturpar o relatório final.

**Etapa 2: ANÁLISE DE ABORDAGENS:** Essa etapa de análise é responsável pela organização das atividades em uma estrutura de análise pré-definida. Nessa etapa é realizada a análise das abordagens em relação ao conteúdo e seus resultados, tais como análises realizadas na RL. Em seguida, analisado o modelo SCNN a ser desenvolvido e realizado experimentação e análises de resultados, assim como definição de bases de dados

a serem utilizadas e parâmetros.

**Etapa 3: DESENVOLVIMENTO, EXPERIMENTAÇÃO E DISCUSSÃO:** Nessa etapa é realizado o desenvolvimento do modelo de SCNN com aprendizagem de poucos disparos para reconhecimento de caracteres a partir de imagens de redações manuscritas de alunos do ensino fundamental público no Brasil. Assim como, é feito experimentos utilizando três bases de dados. Os experimentos são conduzidos utilizando duas funções de perdas distintas, sendo elas, a CrossEntropyLoss e a ContrastiveLoss.

Por fim, é realizada uma discussão acerca dos resultados obtidos nos experimentos e apresentado o modelo com o melhor desempenho com base nas métricas analisadas.

## 1.5 Organização do Documento

O restante deste trabalho está organizado da seguinte forma: no Capítulo 2 aborda a fundamentação teórica, trazendo os conceitos principais relacionados a este estudo. O Capítulo 3 são apresentados os trabalhos relacionados, destacando diversos estudos e abordagens da literatura que têm sido utilizados no reconhecimento de escrita. No Capítulo 4, detalha-se a metodologia utilizada no desenvolvimento da pesquisa. O Capítulo 5 descreve os experimentos realizados e os resultados obtidos de forma detalhada. O Capítulo 6 discute os resultados, suas implicações e contribuições. Por fim, o Capítulo 7 apresenta as conclusões do trabalho.

## 2 Fundamentação Teórica

Este capítulo apresenta conceitos essenciais para a compreensão deste trabalho. Na Seção 2.1 são expostos conceitos iniciais acerca do reconhecimento da escrita, apresentando as etapas do processo de reconhecimento de escrita. Na seção 2.2 trás conceitos acerca das Redes Neurais Convolucionais e no 2.3 conceitos iniciais sobre Few-Shot Learning.

### 2.1 Reconhecimento de Escrita

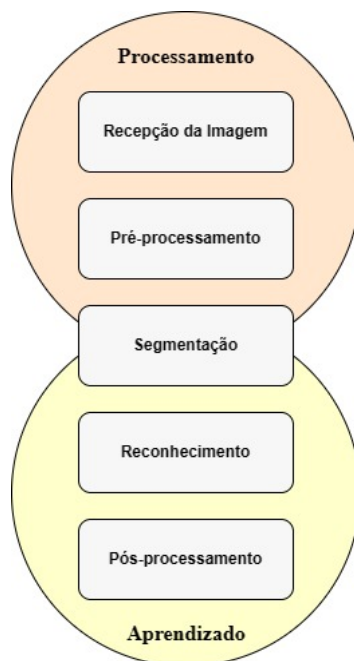
O reconhecimento de escrita ou reconhecimento de textos em imagens contempla por si só um campo de estudo à parte, podendo ser definido como Reconhecimento Óptico de Caracteres (*Optical Character Recognition - OCR*). Este campo, compreende a utilização de alguma ferramenta que identifique, em imagens, figuras que podem ser transformadas em símbolos, e posteriormente reconhecidas como textos para formatos que podem ser compreendidos e manipulados pela máquina (PUROHIT; CHAUHAN, 2016);(CHARLES; KUMAR; SINGH, 2012).

No Reconhecimento de Caracteres Manuscritos (RCM), este processo tem como objetivo identificar e interpretar caracteres alfabéticos, numéricos ou simbólicos escritos à mão. O RCM visa converter informações manuscritas em formato digital para permitir análises automatizadas e resultados mais eficazes e satisfatórios em diversas áreas, assim como na Educação, oferecendo *feedbacks* mais rápidos aos alunos.

O processo de reconhecimento de caracteres manuscritos torna-se uma tarefa complexa em decorrência de diversos aspectos que podem impactar no seu resultado, desde a diversidade na escrita, como na forma de captação dessas imagens, e métodos utilizados para captação, ambiente, entre outros. Como forma de enfrentar esses desafios, proporcionando uma abordagem robusta e adaptativa ao reconhecimento de padrões em dados manuscritos, as Redes Neurais Siameses, uma arquitetura especializada no aprendizado de similaridade entre pares de entradas. Essa abordagem traz, entradas gêmeas, compartilhamento de pesos e camada de distância. O seu treinamento é orientado para aprender representações semelhantes para entradas similares e representações distintas para entradas diferentes.

A execução dessa tarefa pode ocorrer tanto de forma *online*, o texto é reconhecido ao tempo que é escrito, como também de forma *offline*. Conforme a literatura, o processo de reconhecimento de escrita pode ser organizado em cinco etapas, sendo elas: Recepção da Imagem, Pré-processamento, Segmentação, Reconhecimento, e pós-processamento. Como mostra na Figura 2.

Figura 2 – Etapas para o reconhecimento óptico de caracteres.



Fonte: Autoria própria

Esse processo, pode ser resumido basicamente para preparar a imagem para o objetivo final. O grande problema, é que na maioria das vezes o ambiente de captação das imagens não é controlado.

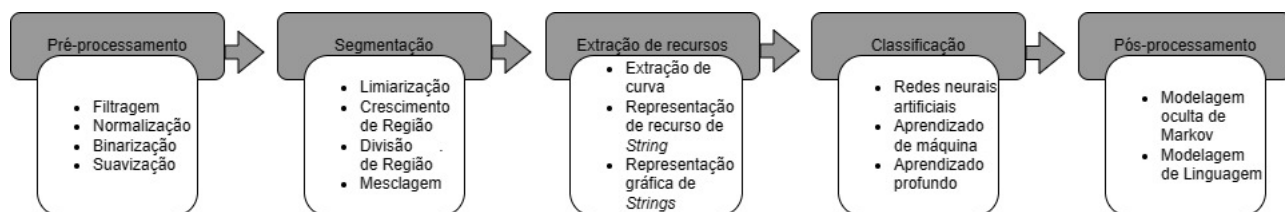
Textos manuscritos podem ser convertidos em formato digital usando duas abordagens distintas. O primeiro é o reconhecimento *off-line*, no qual uma solução depende de um scanner óptico ou de uma câmera digital para digitalizar texto manuscrito em papel (VASHIST; PANDEY; TRIPATHI, 2020). No segundo, a escrita é captada por uma superfície eletrônica (eg, *LCD*) por meio de uma caneta especial (eg, caneta digital), então chamada de reconhecimento *on-line* (Plamondon & Srihari, 2000). O reconhecimento *on-line* geralmente produz melhores resultados porque é uma abordagem simplificada para extrair características relevantes (SEILER; SCHENKEL; EGGIMANN, 1996). No entanto, esta pesquisa se concentra na modalidade de reconhecimento *off-line* porque tal abordagem é mais realista para o contexto de alunos de escolas públicas brasileiras.

No reconhecimento *off-line*, um usuário fornece uma imagem como entrada para o sistema de reconhecimento de manuscrito. Os dados *off-line* consistem em dados de imagem de pixel ordenados no espaço sem a disponibilidade de informações temporais (SEILER; SCHENKEL; EGGIMANN, 1996). O procedimento de reconhecimento envolve a análise da luminância da imagem. Normalmente, a amostragem de dados para capturar um texto manuscrito é de cerca de 300 pontos por polegada. Historicamente, as taxas de reconhecimento relatadas para as taxas de sucesso de reconhecimento *off-line* variam de 78 a 95 por cento (PLAMONDON; SRIHARI, 2000).

Os domínios de aplicação do reconhecimento de escrita *off-line* podem incluir transcrição de documentos, roteamento automático de correspondência e processamento de formulários, cheques e faxes por máquina (SENIOR; ROBINSON, 1998). No entanto, existe uma diferença expressiva entre reconhecer um único caractere ou palavra e reconhecer uma linha de texto ou texto completo. Naturalmente, o último é muito mais desafiador do que o primeiro. Os bons resultados dos métodos ou sistemas de reconhecimento de caracteres e palavras não se dão no reconhecimento de linhas completas de textos manuscritos. Essa dificuldade aumenta ainda mais, quando não há restrição na redação do texto (GRAVES *et al.*, 2009).

(CHOUDHARY *et al.*, 2017) afirmam que o *Processamento Digital de Imagem (PDI)* desempenha um papel vital em um sistema de reconhecimento de escrita. Ele melhora a qualidade de uma imagem, eficiência de armazenamento e qualidade na automação da máquina. Os autores também explicam que o reconhecimento de manuscrito consiste em pré-processamento, segmentação, extração de recursos, classificação e pós-processamento (ver Figura 2).

Figura 3 – Visão geral do processo de reconhecimento de escrita. Adaptado de Choudhary *et al.* (2017).



Fonte: Autoria própria

As atividades de pré-processamento nas imagens de entrada incluem, por exemplo, filtragem, normalização, binarização e suavização. A segmentação de imagem é para simplificar uma imagem subdividindo a imagem em um conjunto de *pixels* com características semelhantes, chamado de segmento. O procedimento de extração de recursos consiste em três métodos principais: extração de curva, representação de recurso de *string* e representação de gráfico de *string*. A tarefa de classificação classifica os novos dados em uma classe ou categoria com base nos dados de treinamento. Por fim, é possível melhorar os resultados do reconhecimento aplicando técnicas de pós-processamento, como o *Hidden Markov Model (HMM)*, e modelando a estrutura da linguagem processada (CHOUDHARY *et al.*, 2017).

(BORA; SHARMA; SONI, 2020) utilizaram uma *Rede Neural Convolutacional (RNC)* e um classificador de código de saída com correção de erros para reconhecimento óptico de caracteres. O treinamento e teste foram realizados no conjunto de dados de imagens de caracteres manuscritos do *NIST* (810.000 caracteres). Da mesma forma, (PAREEK;

SHARMA; SONI, 2020) conduziram o reconhecimento de caracteres guzeráti escritos à mão usando *CNN* e perceptron de múltiplas camadas. Eles utilizaram um conjunto de dados guzeráti com 10.000 imagens. Além disso, (LEE; KIM; LEE, 2021) propuseram um método para reconhecimento de caracteres baseado em imagens de baixa resolução. Focando na educação básica, (ZIN *et al.*, 2021) aplicaram *CNN* para reconhecer caracteres segmentados como parte de uma aplicação para auxiliar na aprendizagem da língua inglesa escrita. Eles testaram o modelo usando mais de 8.000 imagens. No entanto, poucos (ou nenhum) grandes conjuntos de dados fornecem imagens de produções de texto manuscrito, especialmente no contexto brasileiro.

(WU; TAO; BRENNAN, 2016) propuseram um método para a extração de características de imagens, combinando uma *rede neural convolucional* com o método proposto por (GRAY; BRENNAN; TAO, 2007b), chamado de *Feature Fusion Network*. O método consiste em uma *rede neural convolucional* formada por duas partes. A primeira parte contém camadas de convolução e *pooling*, destinadas a extrair características de uma imagem. A segunda parte utiliza uma alteração do método *ELF* desenvolvido por (GRAY; TAO, 2008) para extrair características com base nos canais de cores *RGB*, *YCbCr*, *HSV* e na textura da imagem.

A modificação no método consiste em dividir a imagem de entrada em 16 faixas para que a extração das características seja realizada em cada faixa, produzindo assim um histograma com 16 dimensões para cada canal de cor, que será concatenado no final para formar um único vetor. (WU; TAO; BRENNAN, 2016) denominaram o método modificado de *ELF16*. Após a extração, os vetores gerados em cada uma das etapas são combinados para se obter uma representação mais completa da imagem. Com isso, foi observado que a *Feature Fusion Network* torna a *rede neural convolucional* mais eficaz na extração de características do que o método original desenvolvido por (GRAY; BRENNAN; TAO, 2007b).

(SHAFFI; HAJAMOHIDEEN, 2021) experimentaram com *SCNN* e aprendizado baseado em poucas amostras (*few-shot learning*), para reconhecimento de caracteres escritos à mão em tâmil, evidenciando uma precisão de 83,39%. Por outro lado, (LI; XIE; LU, 2019) utilizaram uma rede Siamesa profunda e aprendizado de um disparo para reconhecimento de caracteres chineses a partir de imagens, relatando resultados promissores. Liu et al. (LIU; WANG; ZHANG, 2022) utilizaram uma rede de similaridade Siamesa e aprendizado de um disparo para reconhecimento de caracteres antigos. No entanto, a literatura carece de estudos focados no reconhecimento de caracteres manuscritos em português, especialmente no contexto de redações de alunos do ensino fundamental no Brasil.

O estudo de abordagens e técnicas utilizadas no reconhecimento de escrita é um tema que vem sendo estudado por diversos pesquisadores. Autores como, (TAPPERT; SUEN; WAKAHARA, 1990) pesquisou o estado da arte do reconhecimento de manuscrito



*online*. Eles examinaram algoritmos de reconhecimento de forma, técnicas de pré e pós-processamento, sistemas experimentais e produtos comerciais. Segundo os autores, o período de desenvolvimento da pesquisa ocorreu em um momento crucial da área, em que eram evidentes renovações de atividades. (SUEN *et al.*, 1993) também analisaram o estado da arte do reconhecimento de escrita. Eles resumem as lições aprendidas, as dificuldades envolvidas e os desafios futuros na área (por exemplo, igualar o desempenho humano). A pesquisa se concentrou no reconhecimento de caracteres isolados e apresentou um breve relato do estado da arte do reconhecimento cursivo.

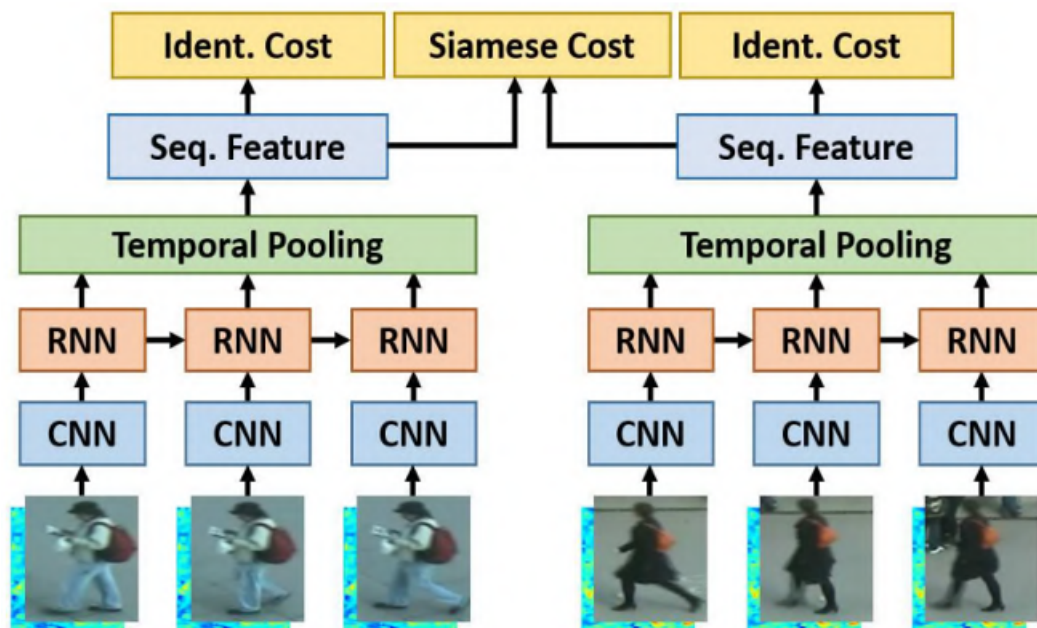
(VASHIST; PANDEY; TRIPATHI, 2020) compararam técnicas de reconhecimento de escrita com foco em aprendizado de máquina. Eles discutem, apresentando prós e contras, o uso de técnicas como máquina de vetores de suporte, k-vizinho mais próximo e redes neurais. Segundo os autores, o reconhecimento de caligrafia a partir de imagens é uma tarefa de classificação desafiadora.

Uma rede neural siamesa foi utilizada por (YI *et al.*, 2014), para encontrar a similaridade entre duas imagens. O modelo proposto consiste em duas redes neurais convolucionais. Cada uma dessas redes convolucionais é composta por duas camadas de convolução, ativadas pela função ReLU, seguidas por camadas de normalização e max pooling, e uma camada totalmente conectada. As sub-redes produzem vetores de tamanho 500 como saída. Para calcular a similaridade entre as imagens, foi utilizada a função cosseno. Os testes realizados nos datasets CUHK03, VIPeR e i-LIDS demonstraram que o método proposto superou os métodos desenvolvidos por (GRAY; BRENNAN; TAO, 2007a) e (ZHAO; OUYANG; WANG, 2013). No entanto, as sub-redes não compartilhavam os mesmos parâmetros, o que resultou em comparações de características menos precisas.

(MCLAUGHLIN; RINCON; MILLER, 2016) desenvolveram uma rede neural siamesa que combina uma rede neural convolucional e uma rede neural recorrente para re-identificação de pessoas com base em vídeo. A rede convolucional gera um mapa de características de cada frame do vídeo, que é então enviado como entrada para a rede recorrente. A particularidade da rede recorrente é sua capacidade de armazenar informações temporais em uma memória durante o processamento. Dessa forma, as informações de cada frame são mantidas e processadas conforme novos mapas de características são obtidos. Uma camada de temporal pooling é utilizada para combinar todas essas características. A função de similaridade utilizada neste método é a Distância Euclidiana. Os datasets utilizados para treinamento foram iLIDS-VID (WANG *et al.*, 2014) e PRID2011 (HIRZER *et al.*, 2011)

Redes Neurais Siamesas (RNS) têm se destacado como uma abordagem eficaz em diversos domínios, especialmente em sistemas de recomendação. Recentemente, (BONNIN *et al.*, 2023) exploraram a integração de SNNs em sistemas de recomendação multimídia, destacando sua capacidade de medir similaridades complexas entre itens e melhorar a

Figura 4 – Arquitetura proposta por (MCLAUGHLIN; RINCON; MILLER, 2016), que consiste em uma rede neural siamesa composta por CNN e RNN.



Fonte: (MCLAUGHLIN; RINCON; MILLER, 2016).

precisão das recomendações. Esta abordagem é particularmente vantajosa em cenários que envolvem dados multimídia, como moda e música, onde a similaridade visual e sonora é crucial. O estudo de Bonnin et al. também enfatiza a importância da escolha adequada das funções de perda, como a função de perda contrastiva e a entropia cruzada binária, que são determinantes para o desempenho do modelo em diferentes aplicações. Além disso, os autores discutem os desafios associados à adaptação das SNNs para diferentes domínios e a necessidade de metodologias de avaliação robustas para garantir a eficácia das recomendações. Estes insights são fundamentais para o desenvolvimento de sistemas de recomendação mais precisos e personalizados, alinhando-se com os objetivos do presente trabalho de aplicar SNNs ao reconhecimento de caracteres manuscritos em contextos educacionais.

## 2.2 Redes Neurais Convolucionais

Uma RNC é caracterizada pela utilização de operações de convolução em pelo menos uma de suas camadas, com o objetivo de aprender padrões de um determinado conjunto de dados. As *CNNs* são especialmente eficazes para o processamento de dados que possuem uma estrutura de grade, como imagens, onde a convolução ajuda a capturar características espaciais e hierárquicas.

As *CNNs* trabalham com dados no formato de múltiplos arrays distribuídos em forma de uma estrutura tridimensional de largura, altura e profundidade. Em imagens

*RGB*, por exemplo, a profundidade da camada de entrada será o número de canais de cor (três, no caso de imagens *RGB*). Conforme os filtros de convolução são aplicados nas camadas de uma *CNN*, as suas dimensões podem ser alteradas, resultando em novas representações das características da imagem.

De acordo com (GEITGEY, 2016), as camadas de uma *Rede Neural Convolutional (CNN)* seguem diferentes etapas de refinamento. Primeiramente, uma camada de convolução, que funciona como uma janela deslizante, desloca-se sobre a imagem em passos específicos (*stride*). Essa camada secciona a imagem em um conjunto de sub-regiões sobrepostas, nas quais são aplicados filtros matriciais que extraem características relevantes da imagem.

Como descrito por (SERMANET; LECUN, 2011), o processo de convolução ocorre em três estágios. No primeiro estágio, operações lineares de convolução extraem informações representadas por valores positivos e negativos. No segundo estágio, esses valores são transformados em não-lineares através da função de ativação *ReLU (Rectified Linear Unit)*, preservando apenas os valores positivos. No terceiro estágio, a etapa de *pooling* simplifica a imagem resultante, preservando de forma seletiva apenas a informação essencial. Esse processo gera mapas de convolução, que são posteriormente fornecidos como entrada a um classificador.

### 2.2.1 Redes Neurais Convolucionais Siamesas

Conforme descrito por Bromley et al. (1994), uma rede neural siamesa é composta por duas sub-redes idênticas que são conectadas em suas saídas. Esta arquitetura foi inicialmente desenvolvida para buscar correspondências em imagens de assinaturas. Cada sub-rede processa uma entrada diferente e a transforma em um descritor de características. Em seguida, os dois descritores gerados são comparados para determinar a semelhança entre eles, resultando na saída da rede. É fundamental que as sub-redes compartilhem os mesmos parâmetros e pesos, garantindo que suas saídas sejam comparáveis.

As *Siamese Neural Networks (SNNs)* surgiram em 1994 como uma arquitetura de rede neural artificial onde duas redes neurais idênticas são utilizadas para calcular a similaridade entre dois elementos. Esse tipo de arquitetura é particularmente adequado para situações em que aprender uma medida de similaridade é fundamental para a aplicação. Além disso, as *SNNs* têm se mostrado bastante escalonáveis e eficientes.

Cada sub-rede na *SNN* processa uma entrada diferente, gerando um descritor de características. Esses descritores são então comparados para determinar a semelhança entre as entradas. É crucial que as sub-redes compartilhem os mesmos parâmetros e pesos para garantir que as saídas sejam comparáveis. A arquitetura foi inicialmente apresentada para buscar correspondências em imagens de assinaturas, mas desde então tem sido aplicada

em uma ampla variedade de tarefas, incluindo reconhecimento facial e verificação de assinaturas (BROMLEY *et al.*, 1994a).

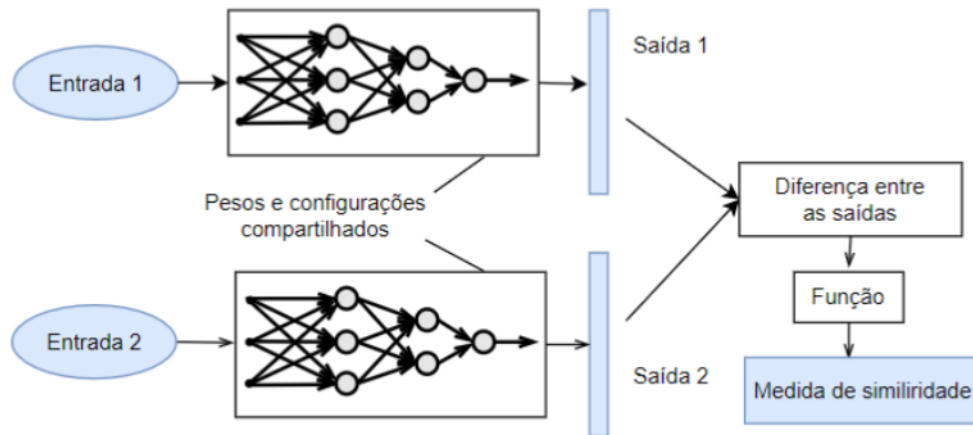


Figura 5 – Arquitetura simplificada de uma rede siamesa (Fonte: (CHAVES *et al.*, 2021)).

Existem duas funções de perda principais usadas em modelos de Redes Neurais Siamesas: a Entropia Cruzada Binária e a Perda Contrastiva.

A *Binary Cross Entropy* é utilizada para determinar se dois elementos pertencem à mesma classe ou a classes diferentes. A fórmula é expressa da seguinte maneira:

$$\text{Perda} = - [Y \log(Y_{\text{pred}}) + (1 - Y) \log(1 - Y_{\text{pred}})] \quad (2.1)$$

onde:

- $Y$  é o valor do rótulo, sendo 1 se ambos os pares pertencem à mesma classe e 0 caso contrário.
- $Y_{\text{pred}}$  é o valor do rótulo previsto pela rede siamesa.

#### Perda Contrastiva

A Perda Contrastiva é mais adequada ao problema abordado pelas Redes Siamesas, pois o objetivo é diferenciar (e não classificar) entre dois elementos. Essa função é utilizada para medir a capacidade da rede de encontrar as semelhanças entre as imagens. Assim, essa função deve aprender os parâmetros  $W$ , de forma que os exemplos mais semelhantes fiquem mais próximos e os mais diferentes sejam separados. Para garantir que isso ocorra, durante o treinamento, cada par de imagens recebe um rótulo  $Y$ , sendo  $Y = 0$  quando são semelhantes, ou seja, de um mesmo grupo, e  $Y = 1$  quando são diferentes. A fórmula da Perda Contrastiva é expressa como:

$$L(W, Y, X_1, X_2) = (1 - Y) \frac{1}{2} (D_W)^2 + Y \frac{1}{2} \{\max(0, m - D_W)\}^2 \quad (2.2)$$

onde:

- $Y$  é o valor do rótulo, sendo 1 se os elementos são diferentes e 0 caso contrário.
- $D_W$  é a distância Euclidiana entre os vetores de características das duas entradas.
- $m$  é a margem, um valor hiperparamétrico que define a distância mínima que vetores de características de elementos de diferentes classes devem ter.

onde  $m > 0$  é a margem, tal que  $(X_1, X_2)$  seja um par de entradas positivas e  $(X_1, X'_2)$  seja um par negativo. Temos que  $D_W(X_1, X_2) + m < D_W(X_1, X'_2)$  (HADSELL; CHOPRA; LECUN, 2006).

## 2.3 Few-Shot Learning

O objetivo principal do aprendizado de máquina é aprender adquirindo novas fontes externas de conhecimentos, na busca de se tornar mais eficiente. No entanto, um dos desafios significativos enfrentados nessa área é a dependência de grandes quantidades de dados rotulados para treinar modelos robustos. Em muitos cenários do mundo real, a obtenção de grandes volumes de dados anotados é inviável devido a custos, tempo ou recursos limitados. Para resolver esse problema, técnicas de aprendizado foram desenvolvidas para permitir que modelos de aprendizado de máquina funcionem eficientemente com conjuntos de dados reduzidos.

*Few-Shot Learning (FSL)* é uma dessas técnicas inovadoras, focada no treinamento de modelos com uma quantidade mínima de dados de treinamento. O *FSL* visa aperfeiçoar modelos a generalizarem a partir de um pequeno número de exemplos por classe, abordando a limitação de dados de maneira eficaz. Esse paradigma é particularmente relevante em situações onde a coleta de dados é difícil ou quando novas classes de dados aparecem constantemente, exigindo que o modelo se adapte rapidamente.

Existem diversas abordagens para implementar *Few-Shot Learning*, cada uma explorando diferentes estratégias para lidar com a escassez de dados. Uma delas é o *Aprendizado Baseado em Métricas*, que envolve a aprendizagem de uma função de similaridade ou distância que pode ser usada para comparar novos exemplos com um pequeno conjunto de exemplos rotulados. Redes siamesas e redes de protótipos são exemplos comuns dessa categoria.

O *Aprendizado de Transferência* ou *Transfer Learning* é utilizado para adaptar modelos pré-treinados em grandes conjuntos de dados para novas tarefas com poucos

dados. Técnicas como *fine-tuning* e *feature extraction* permitem que o modelo aproveite o conhecimento adquirido anteriormente. (HOWARD; RUDER, 2018) demonstraram o sucesso do *Transfer Learning* em modelos de linguagem com o *ULMFiT*.

Os *Métodos de Meta-Aprendizado*, também conhecidos como "aprendizado de aprender", envolvem o treinamento de um modelo em uma variedade de tarefas para que ele possa rapidamente se adaptar a novas tarefas com poucos exemplos. Algoritmos como *MAML (Model-Agnostic Meta-Learning)*, proposto por (FINN; ABBEEL; LEVINE, 2017), são exemplos dessa abordagem.

A Implementação do *FSL* no Modelo Siamesa, ocorre de forma que o *FSL* é implementado utilizando uma *Rede Neural Convolutiva Siamesa (RNCS)*. A *SCNN* é composta por duas sub-redes idênticas que compartilham os mesmos pesos e parâmetros, permitindo a comparação de pares de entradas para determinar a similaridade entre elas. Essa abordagem é eficaz para aprender representações discriminativas com poucos exemplos de cada classe. A utilização de redes siamesas para tarefas de *Few-Shot Learning* foi popularizada por (KOCH; ZEMEL; SALAKHUTDINOV, 2015), que demonstraram sua eficácia em tarefas de reconhecimento de imagens.

### 2.3.1 Vantagens e Desafios do Few-Shot Learning

A utilização do *FSL* traz vantagens como a eficiência de dados, permitindo que modelos aprendam de um pequeno número de exemplos, reduzindo a necessidade de grandes volumes de dados rotulados. Outra vantagem seria a adaptabilidade, facilitando a adaptação rápida a novas classes ou tarefas, essencial em ambientes dinâmicos, como também a redução de custos, pois diminui os custos associados à coleta e anotação de dados.

Assim como as vantagens, a sua implementação traz desafios relacionados à complexidade computacional, pois algumas abordagens, como o meta-aprendizado, podem ser computacionalmente intensivas. Os ajustes de hiperparâmetros ótimos podem ser desafiantes devido à quantidade limitada de dados, e a dependência de modelos pré-treinados, pois o desempenho pode depender da qualidade desse modelo.

### 2.3.2 Aplicações do FSL

O *FSL* tem sido aplicado com sucesso em diversas áreas, tais como no reconhecimento de imagens: classificação de objetos e identificação de faces com poucos exemplos. (VINYALS *et al.*, 2016) propuseram a rede de correspondência (*Matching Network*) para tarefas de *Few-Shot Learning*, mostrando resultados promissores em reconhecimento de imagens. No processamento de linguagem natural: reconhecimento de entidades e classificação de texto em cenários com poucos dados anotados. Na robótica: adaptação rápida a

novas tarefas e ambientes com treinamento limitado e no reconhecimento de caracteres.

## 3 Trabalhos Relacionados

Neste capítulo são apresentados alguns trabalhos disponíveis na literatura que tratam do problema de reconhecimento de caracteres manuscritos. Considerando que o processo de reconhecimento de caracteres manuscritos envolve determinar se um caractere em uma imagem corresponde ao mesmo caractere que aparece em outra imagem, o desafio central reside na variabilidade de estilos de escrita, caligrafia e condições de iluminação. Sendo assim, a legibilidade impacta significativamente a precisão do reconhecimento, pois caracteres claramente legíveis facilitam a extração de características relevantes e a correspondência precisa entre eles.

A seguir, é apresentada uma revisão de trabalhos relevantes na área, destacando as técnicas utilizadas e as bases de dados empregadas

(GOYANI; DANI; DIXIT, 2018) em seu estudo, trás uma pesquisa abrangente sobre métodos de reconhecimento de caracteres manuscritos, discutindo tanto técnicas tradicionais quanto modernas. A pesquisa destaca a importância da legibilidade dos caracteres na melhoria da precisão do reconhecimento e examina várias abordagens, incluindo aprendizado profundo e métodos baseados em redes neurais. O estudo também enfatiza os desafios no reconhecimento de caracteres manuscritos e os avanços feitos nos últimos anos para enfrentar essas questões. A base de dados utilizada é variada, cobrindo um amplo espectro de conjuntos de dados gerais.

jà (RAJASEKAR *et al.*, 2022), conduz um estudo que propõe uma metodologia baseada em Redes Neurais Convolucionais (CRNCs) para reconhecer eficientemente caracteres manuscritos. O foco está no impacto da legibilidade dos caracteres na precisão do modelo. O estudo explora diferentes opções de design para arquiteturas de CNN, como número de camadas, tamanho do passo e tamanho do kernel, para otimizar o desempenho dos sistemas de reconhecimento de caracteres manuscritos. A base de dados utilizada é variada, abrangendo diferentes conjuntos de dados gerais.

(DEORE; PRAVIN, 2019), em seu trabalho, tem como foco o reconhecimento de caracteres manuscritos Devanagari, utilizando Redes Neurais Convolucionais Profundas (RNCP) ajustadas. O estudo enfatiza o papel da legibilidade dos caracteres na precisão do reconhecimento e combina técnicas tradicionais de extração de características como *Histogram of Oriented Gradients (HOG)*, *Local Binary Patterns (LBP)* e *Zone-Based Features*, com abordagens modernas de aprendizado profundo, como a utilização de camadas convolucionais profundas e técnicas de regularização, para aprimorar o desempenho do sistema de reconhecimento. A base de dados utilizada é específica para o conjunto de dados *Devanagari*.



O estudo conduzido por (BALDOMINOS; SAEZ; ISASI, 2020), revisa métodos de ponta para reconhecimento de dígitos manuscritos usando os conjuntos de dados MNIST e EMNIST. A pesquisa discute a importância da legibilidade dos caracteres para alcançar alta precisão de reconhecimento e explora várias técnicas de aprendizado de máquina e aprendizado profundo. O estudo fornece uma comparação detalhada de diferentes métodos e destaca os avanços na área. A pesquisa destaca pontos como, técnicas utilizadas com diversos métodos de aprendizado de máquina, incluindo Support Vector Machines (SVM), Random Forests, K-Nearest Neighbors (KNN) e abordagens baseadas em Redes Neurais Convolucionais (RNCs). A análise comparativa mostrou que métodos baseados em CNNs superaram consistentemente outras técnicas em termos de precisão, especialmente quando treinados com um grande número de exemplos. E destaca os avanços no uso de aprendizado profundo, como a introdução de arquiteturas mais complexas, assim como, técnicas de regularização que ajudam a melhorar a generalização dos modelos.

Já (ALOM *et al.*, 2018) aborda o reconhecimento de caracteres manuscritos Bangla utilizando Redes Neurais Convolucionais Profundas de última geração (RNCP). O estudo foca em como a legibilidade dos caracteres afeta a precisão do reconhecimento e apresenta ajustes específicos para melhorar a eficácia do sistema de reconhecimento de caracteres Bangla. A base de dados utilizada é específica para caracteres Bangla. Ou seja, (ALOM *et al.*, 2018) em seu trabalho, definiu que a legibilidade dos caracteres afeta diretamente a precisão do reconhecimento ao demonstrar que caracteres mais claramente legíveis facilitam a extração de características relevantes, resultando em maior acurácia na classificação. Concluiu-se que, ao melhorar a legibilidade dos caracteres através de pré-processamento e técnicas de aumento de dados, a precisão do modelo de reconhecimento de caracteres Bangla aumentou significativamente.

Os trabalhos discutidos neste capítulo fornecem uma base sólida de conhecimento sobre as técnicas e desafios no reconhecimento de caracteres manuscritos. A comparação entre diferentes abordagens e bases de dados utilizadas destaca a diversidade de métodos disponíveis e a importância da legibilidade dos caracteres na precisão dos modelos de reconhecimento. O estudo desenvolvido neste trabalho se alinha com essas pesquisas, contribuindo para o avanço da tecnologia de reconhecimento de caracteres manuscritos, especialmente em contextos educacionais.

(MUKHERJEE, 2024), em seu estudo, propôs uma abordagem holística baseada em redes neurais profundas para o reconhecimento óptico de caracteres (ROC) de documentos manuscritos, combinando redes neurais convolucionais (RNC) e redes longas de memória de curto prazo bidirecionais (BLSTMs). O objetivo do estudo foi desenvolver um modelo híbrido que extraísse características espaciais e temporais de imagens de palavras manuscritas, melhorando a precisão do reconhecimento. O modelo foi treinado e testado em três bases de dados públicas (CMATERdb2.1.2, IIIT-HW-Dev, IIIT-HW-Telugu) e

alcançou precisão de 96,42%, 94,79% e 95,07% para as bases de dados Bengali, Devanagari e Telugu, respectivamente. A pesquisa não aborda diretamente a questão da legibilidade dos caracteres, mas a alta precisão sugere que o modelo é eficaz em lidar com variabilidade na escrita.

(MISHRA ATMAN E RAM, 2023), também utilizaram em seu estudo, redes neurais convolucionais customizadas (RNCC) com pooling, regularização por dropout e entropia cruzada categórica para o reconhecimento de textos manuscritos. O estudo foi conduzido utilizando o conjunto de dados NIST, que contém 101,784 imagens em 47 classes. O objetivo foi criar um modelo eficiente para o reconhecimento de caracteres manuscritos com uma precisão considerável, embora o artigo não mencione explicitamente a questão da legibilidade dos caracteres.

Em um estudo realizado por (SHARMA REYA E KAUSHIK, 2020), foi realizado uma revisão abrangente sobre o reconhecimento offline de scripts manuscritos Indicos, explorando o uso de CNNs e LSTMs. O objetivo do estudo foi fornecer uma visão geral das técnicas de ponta utilizadas para o reconhecimento de scripts Indicos, avaliando a eficácia das redes neurais em diversos datasets de scripts Indicos.

(CHAMMAS EDGARD E MOKBEL, 2023), desenvolveu um modelo de rede neural recorrente convolucional profunda (CRNN) para o reconhecimento de documentos históricos manuscritos com poucos dados rotulados. O estudo utilizou o conjunto de dados READ, focando na eficácia de redes neurais profundas em cenários com dados limitados. A pesquisa destaca a importância da precisão no reconhecimento de documentos históricos.

E por fim, (JUNGO MICHAEL E WOLF, 2023), propuseram uma abordagem baseada em Transformers e LSTMs, inspirada no algoritmo k-means, para a segmentação de caracteres manuscritos online. O estudo utilizou os conjuntos de dados IAM-OnDB e HANDS-VNOnDB, com o objetivo de melhorar a segmentação de caracteres manuscritos ao atribuir clusters de caracteres com base em consultas aprendidas no bloco decodificador Transformer.

Este capítulo traz trabalhos que abrangem uma variedade de abordagens para o reconhecimento de caracteres manuscritos. Esses estudos fornecem uma base sólida para entender as diversas técnicas aplicadas ao reconhecimento de caracteres manuscritos e sua eficácia em diferentes contextos e bases de dados.

Este estudo se destaca pelo uso da base de dados brasileira, um avanço significativo na área de reconhecimento de caracteres manuscritos, dado que, muitos estudos anteriores se concentram em bases de dados estrangeiras. Além disso, foi utilizada a técnica de few-shot learning, que permite ao modelo aprender de forma eficaz com uma quantidade limitada de exemplos anotados. Essa abordagem é particularmente benéfica para situações onde os dados são escassos, garantindo uma boa precisão em contextos específicos do

Brasil.

A Tabela 1 apresenta uma visão geral das diversas áreas e aplicações onde o modelo de Rede Neural Siamesa (SCNN) pode ser utilizado, juntamente com os objetivos dos estudos relevante. Esses estudos destacam a versatilidade e a eficácia das redes siamesas em diversas aplicações, demonstrando seu potencial para resolver problemas complexos de reconhecimento de padrões em diferentes setores.

Tabela 1 – Modelo Siamesa em Diferentes Áreas e Objetivos

Área	Aplicação	Referências e Objetivos
Reconhecimento de Caracteres Manuscritos	Processamento de Formulários e Documentos	(SMITH et al., 2020): Desenvolver e avaliar um sistema de reconhecimento de caracteres manuscritos para processamento de formulários e documentos. (ZHAO et al., 2019): Apresentar um novo método de reconhecimento de caracteres manuscritos que utiliza técnicas de aprendizado profundo para aumentar a taxa de acerto na digitalização de documentos manuscritos.
Reconhecimento de Caracteres Manuscritos	Reconhecimento de Assinaturas	(LIU et al., 2021): Desenvolver um sistema robusto para reconhecimento de assinaturas manuscritas, utilizando redes neurais siamesas para diferenciar entre assinaturas verdadeiras e falsas. (KUMAR et al., 2018): Propor um método eficiente de verificação de assinaturas que combina técnicas de processamento de imagem e aprendizado de máquina para autenticação de documentos.
		Continua na próxima página

Tabela 1 continuada da página anterior

Área	Aplicação	Referências e Objetivos
Saúde	Diagnóstico de Imagens Médicas	<p>(BROWN et al., 2020): Aplicar técnicas de aprendizado profundo no diagnóstico de imagens médicas, melhorando a detecção e classificação de doenças através da análise automatizada de imagens.</p> <p>(CHEN et al., 2019): Desenvolver um sistema de apoio ao diagnóstico médico que utiliza redes neurais convolucionais para analisar imagens médicas e auxiliar médicos na identificação de patologias.</p>
Saúde	Análise Histopatológica	<p>(NGUYEN et al., 2021): Implementar um modelo de aprendizado profundo para a análise histopatológica, com foco na detecção e classificação de tecidos cancerígenos em amostras biológicas.</p> <p>(SHARMA et al., 2018): Utilizar redes neurais para automatizar a análise de lâminas histológicas, aumentando a precisão e a velocidade do diagnóstico histopatológico.</p>
Continua na próxima página		

Tabela 1 continuada da página anterior

Área	Aplicação	Referências e Objetivos
Segurança e Vigilância	Reconhecimento Facial	<p>(WANG et al., 2020): Desenvolver um sistema de reconhecimento facial utilizando redes neurais profundas para melhorar a segurança e a vigilância em ambientes públicos.</p> <p>(PATEL et al., 2019): Propor uma abordagem para reconhecimento facial em tempo real, visando aumentar a acurácia e a rapidez na identificação de indivíduos em grandes multidões.</p>
Segurança e Vigilância	Detecção de Anomalias em Vídeos	<p>(SINGH et al., 2021): Desenvolver um sistema para detecção de anomalias em vídeos de vigilância, utilizando aprendizado de máquina para identificar atividades suspeitas em tempo real.</p> <p>(GOMEZ et al., 2018): Aplicar técnicas de aprendizado profundo na análise de vídeos de segurança, focando na identificação automática de comportamentos anômalos.</p>
Continua na próxima página		

Tabela 1 continuada da página anterior

Área	Aplicação	Referências e Objetivos
Educação	Correção Automática de Provas e Trabalhos	<p>(LEE et al., 2020): Desenvolver um sistema automatizado para correção de provas e trabalhos manuscritos, utilizando reconhecimento de caracteres para facilitar o trabalho dos educadores.</p> <p>(WANG et al., 2019): Propor um método de correção automática de exercícios escolares, combinando técnicas de reconhecimento de escrita com análise de conteúdo.</p>
Educação	Personalização do Ensino	<p>(JOHNSON et al., 2021): Criar um sistema de ensino personalizado que utilize análise de dados de desempenho estudantil para adaptar o material didático às necessidades individuais dos alunos.</p> <p>(GARCIA et al., 2018): Desenvolver uma plataforma de ensino que personaliza a experiência de aprendizado com base na análise do progresso e dificuldades de cada aluno.</p>
Continua na próxima página		

Tabela 1 continuada da página anterior

Área	Aplicação	Referências e Objetivos
Indústria Financeira	Verificação de Documentos	(KUMAR et al., 2020): Implementar um sistema de verificação de documentos financeiros que utiliza reconhecimento de caracteres para automatizar processos de validação e autenticação. (KIM et al., 2019): Propor uma abordagem para verificação automática de documentos financeiros, utilizando aprendizado de máquina para aumentar a precisão e a segurança.
Indústria Financeira	Deteção de Fraudes	(ZHANG et al., 2021): Desenvolver um sistema de deteção de fraudes em transações financeiras, aplicando técnicas de aprendizado profundo para identificar padrões suspeitos. (KHAN et al., 2018): Criar um modelo de aprendizado de máquina para deteção e prevenção de fraudes, utilizando análise de dados em tempo real para identificar atividades fraudulentas.

Fonte: Autoria Própria.

## 4 Materiais e Métodos

Este capítulo apresenta a metodologia adotada nesse trabalho para o desenvolvimento, implementação e experimentação do modelo de proposto.

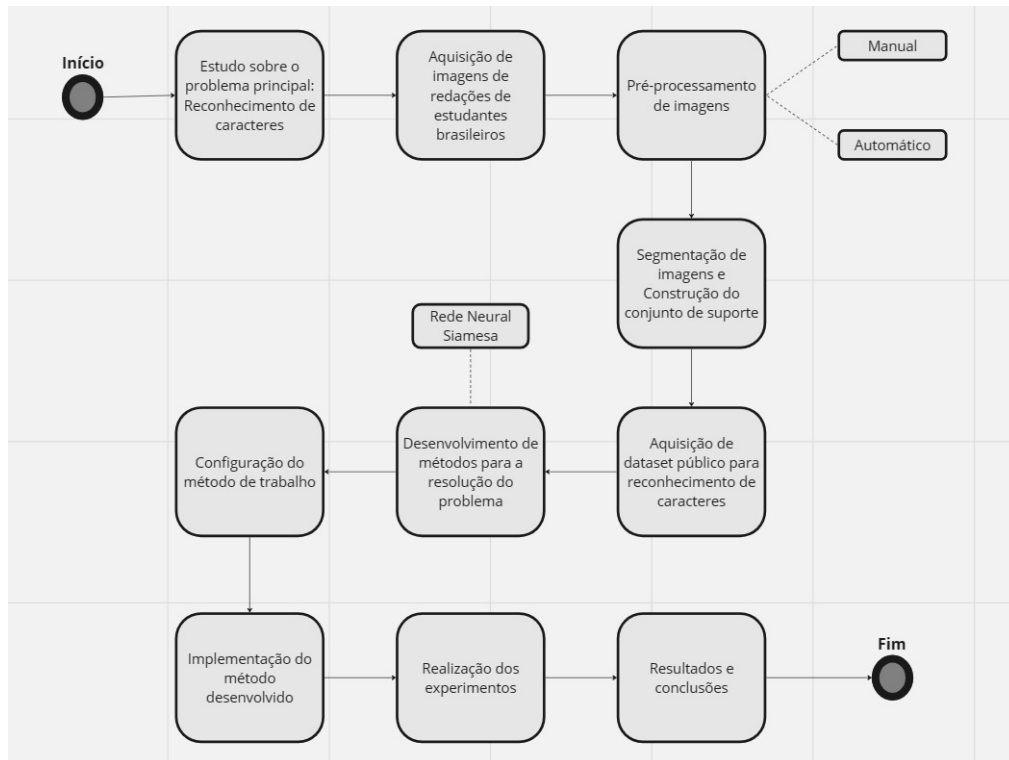


Figura 6 – Passos metodológicos para implementar e validar o modelo.

### 4.1 Problema Principal

A fase inicial envolveu a compreensão do problema de reconhecimento de caracteres. Assim, foi abordado questões fundamentais e abordagens existentes. No primeiro passo, realizou-se um estudo sobre o reconhecimento de caracteres manuscritos. Considerando técnicas e abordagens utilizadas por outros autores, explorando pesquisas anteriores (por exemplo, (SUWANWIWAT; PAL; BLUMENSTEIN, 2016; GRAVES *et al.*, 2009; FALK *et al.*, 2011; LIN *et al.*, 2020)) e métodos existentes para entender os desafios e soluções propostas nesse campo.

### 4.2 Coleta de Dados e Pré-processamento

Neste estudo, foram coletadas 440 imagens de redações manuscritas de estudantes brasileiros. Essas imagens passaram por um processo de pré-processamento que incluiu



métodos manuais e automáticos para melhorar a qualidade e legibilidade das imagens antes da segmentação.

Inicialmente, foram aplicadas técnicas automáticas de pré-processamento utilizando a Biblioteca de Visão Computacional de Código Aberto (*OpenCV*). Essas técnicas incluíram:

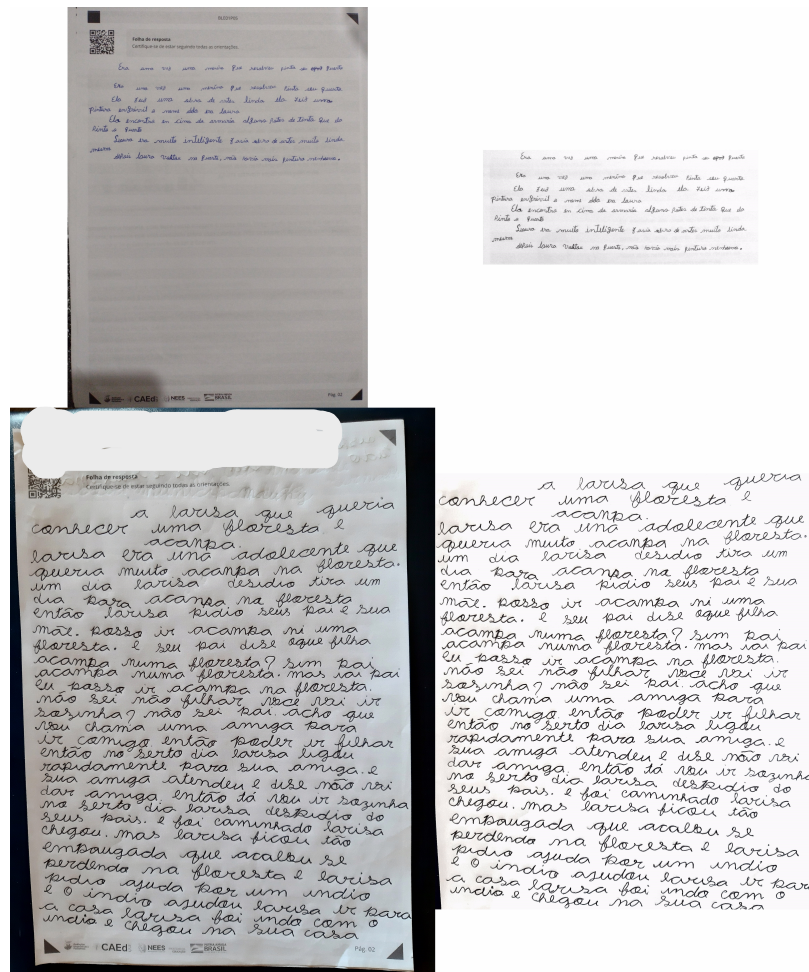
- **Conversão de escala de cinza:** Transformar a imagem colorida em escala de cinza para simplificar o processamento.
- **Desfoque Gaussiano:** Suavizar a imagem para reduzir o ruído e os detalhes finos.
- **Binarização:** Converter a imagem em preto e branco, onde os pixels brancos representam o texto e os pixels pretos representam o fundo.
- **Remoção de ruído:** Filtrar pequenos elementos indesejados para melhorar a clareza do texto.

Essas etapas automáticas são cruciais para tratar características gerais das imagens e preparar os dados para as fases subsequentes de processamento. No entanto, mesmo após essas melhorias, foi necessário realizar ajustes manuais adicionais para aprimorar ainda mais a qualidade e legibilidade das imagens. Esses ajustes manuais incluíram:

- **Edição de imagem:** Correção de distorções e ruídos não eliminados automaticamente.
- **Recorte:** Remoção de áreas desnecessárias ao redor do texto para focar apenas nas regiões de interesse.
- **Ajuste de brilho e contraste:** Melhoria da visibilidade dos caracteres ajustando os níveis de brilho e contraste das imagens.

A Figura 7 ilustra exemplos do pré-processamento manual e automático realizado.

Figura 7 – Pré-processamento manual e automático.



A Figura 7 apresenta exemplos de imagens submetidas a diferentes processos de pré-processamento. À direita, estão as imagens após o pré-processamento automático, enquanto à esquerda são mostradas as imagens que passaram por técnicas de pré-processamento manual.

**Imagens à direita (Pré-processamento automático):** Essas imagens foram submetidas a processos automáticos de pré-processamento que incluem ajustes de brilho, contraste e nitidez, além de possíveis correções de distorções e alinhamento de texto. O objetivo dessas técnicas é uniformizar as características visuais das imagens para facilitar o reconhecimento automático de caracteres. O pré-processamento automático é eficiente em lidar com grandes volumes de dados, porém, pode não capturar todas as nuances da escrita manuscrita que são cruciais para uma interpretação precisa.

**Imagens à esquerda (Pré-processamento manual):** As imagens à direita foram ajustadas utilizando técnicas manuais, onde intervenções humanas foram feitas para melhorar a qualidade e a legibilidade do texto. Isso pode incluir a remoção de artefatos indesejados, aprimoramento das bordas dos caracteres e correção de erros não detectados pelos métodos automáticos. O pré-processamento manual, embora mais demorado, pode

resultar em uma qualidade superior de dados, especialmente em situações onde os métodos automáticos falham em capturar todos os detalhes necessários para um reconhecimento preciso.

### 4.2.1 Segmentação e construção de conjunto de suporte

A segmentação de caracteres foi realizada utilizando algoritmos automáticos complementados por intervenções manuais, quando necessário, para extrair caracteres individuais para análise e treinamento. Utilizou-se a biblioteca *OpenCV* e funções adicionais das bibliotecas padrão do *Python* para este processo. As principais funções utilizadas incluíram:

- `cv2.imread`: Carregar a imagem em memória.
- `cv2.cvtColor`: Converter a imagem para escala de cinza.
- `cv2.GaussianBlur`: Aplicar desfoque gaussiano para suavização.
- `cv2.threshold`: Realizar a binarização da imagem.
- `cv2.bitwise_not`: Inverter as cores da imagem para destacar o texto.
- `cv2.findContours`: Encontrar contornos dos caracteres para segmentação.

A biblioteca *NumPy* também foi utilizada para manipulações adicionais, como a criação de matrizes para máscaras usadas na inversão de cores com a função `np.zeros_like`. A Figura 8 apresenta exemplos de imagens segmentadas.

O algoritmo de segmentação de caracteres envolve várias etapas detalhadas:

1. **Conversão para escala de cinza e desfoque gaussiano**: Estas etapas suavizam a imagem e reduzem o impacto do ruído, preparando a imagem para a binarização.
2. **Binarização**: Segmenta a imagem em regiões de interesse, destacando o texto em branco sobre um fundo preto.
3. **Inversão de cores**: Realça o texto, invertendo as cores para facilitar a detecção dos contornos.
4. **Detecção e recorte de contornos**: Identifica as regiões dos caracteres e as recorta para criar imagens individuais de cada caractere.

A Figura 8 ilustra o resultado do processo de segmentação das imagens, onde cada caractere manuscrito foi isolado de sua fonte original para facilitar a análise e o reconhecimento automático. O processo de segmentação é fundamental para separar e identificar individualmente cada letra, garantindo que o modelo de reconhecimento possa

focar nas características específicas de cada caractere. As letras segmentadas exibidas na figura mostram a diversidade de formas e estilos de escrita capturados no conjunto de dados, desde variações na espessura dos traços até diferenças na curvatura e ângulo das letras. Este tipo de segmentação é crucial para treinar modelos robustos que possam lidar com a grande variabilidade presente na escrita manuscrita de diferentes indivíduos.

Figura 8 – Exemplos de imagens segmentadas.



Após a segmentação, foi necessária uma seleção manual rigorosa das imagens. Das 134.603 imagens segmentadas, 117.999 foram descartadas por não atenderem aos critérios de qualidade estabelecidos, como:

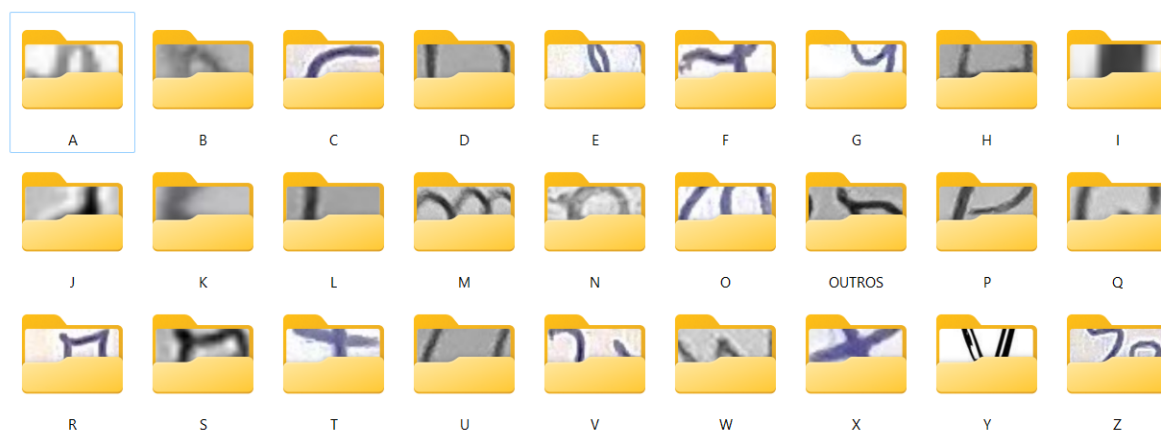
- Caligrafia inadequada: Consideramos caligrafia inadequada quando os caracteres são ilegíveis, excessivamente distorcidos, ou apresentam inconsistências significativas na sua formação, dificultando tanto o reconhecimento automático quanto a análise manual. Isso inclui letras com formas pouco definidas, onde o modelo pode ter dificuldade em identificar a classe correta.
- Espaçamento incorreto entre letras e palavras:\*\* Imagens onde o espaçamento entre letras ou palavras é irregular foram descartadas, pois isso pode causar problemas no reconhecimento dos caracteres e na segmentação. Espaçamento inconsistente pode levar a erros na interpretação do início e término das palavras, além de confundir o modelo quanto à estrutura da escrita.
- Presença de rasuras: Imagens contendo rasuras, que são alterações ou correções feitas sobre o texto original, foram excluídas. Rasuras podem introduzir ruídos visuais que distorcem a forma original das letras, dificultando a leitura automática e comprometendo a precisão do reconhecimento de caracteres.
- Imagens contendo múltiplos caracteres: Para garantir que cada imagem representasse um único caractere, foram descartadas aquelas que continham mais de um caractere.

A presença de múltiplos caracteres em uma única imagem pode confundir o modelo, comprometendo a eficácia do processo de segmentação e reconhecimento.

- Ruídos e bordas indesejadas: Imagens com ruídos (como manchas, linhas ou sombras) ou com bordas indesejadas (áreas fora do campo do caractere, como margens de cadernos ou outras partes da folha) foram eliminadas. Esses elementos podem interferir na qualidade da imagem e dificultar a extração precisa das características dos caracteres pelo modelo de reconhecimento.

Ao final, 16.605 imagens foram selecionadas manualmente, organizadas em pastas correspondentes a cada letra específica, garantindo uma distribuição balanceada de amostras em cada classe.

Figura 9 – Conjunto de suporte.



A criação do conjunto de suporte é uma fase crítica. O foco é compilar um conjunto diversificado de imagens segmentadas representando cada letra do alfabeto. Isso envolve a identificação e escolha cuidadosa de imagens representativas para cada letra, levando em consideração as variações na escrita, estilos de caligrafia e condições de iluminação. A organização do conjunto de suporte, com imagens separadas em pastas distintas para cada letra, e a garantia de uma distribuição balanceada de amostras em cada classe são fundamentais para o sucesso do treinamento do modelo.

A importância de ter amostras balanceadas nas classes é fundamental no contexto do *few-shot learning*. Em situações onde há poucas amostras disponíveis para treino, um conjunto de suporte balanceado assegura que o modelo aprenda de maneira equitativa sobre todas as classes. Se algumas classes estiverem sub-representadas, o modelo poderá se tornar enviesado, resultando em um desempenho inferior para aquelas classes. A distribuição balanceada permite que o modelo generalize melhor, melhorando a precisão e a robustez na identificação de cada letra, independentemente da quantidade limitada de dados disponíveis para treinamento.

Além disso, no conjunto de suporte formado, é essencial incluir vários estilos de escrita de caracteres. Isso abrange diferentes caligrafias, tamanhos e formatos de letras, bem como variações introduzidas por diferentes condições de iluminação e qualidade das imagens. Ao incorporar essa diversidade, o modelo se torna mais resiliente e capaz de lidar com a variabilidade encontrada em situações reais. A inclusão de múltiplos estilos garante que o sistema de reconhecimento possa identificar corretamente os caracteres, independentemente das diferenças individuais na escrita, aumentando assim a precisão e a eficácia do modelo no reconhecimento de texto manuscrito.

### 4.3 Aquisição da base de dados

Nesta etapa, foi incorporado três conjuntos de dados públicos amplamente reconhecidos na comunidade de aprendizado de máquina para o treinamento da *SCNN*: EMNIST, NIST Special Database 19 e *IMAGENET* (ver Tabela 2).

Tabela 2 – Conjuntos de Dados Utilizados nos Experimentos

Conjunto de Dados	Descrição	Link de Acesso
EMNIST	Consiste em mais de 800 mil imagens de caracteres manuscritos em diferentes estilos. Inclui uma variedade rica de caracteres alfabéticos e numéricos.	EMNIST Dataset
NIST <i>Special Database 19</i>	Contém formas e caracteres manuscritos, especializado em formas e caracteres escritos à mão. Fornece uma base rica e variada de caracteres, permitindo avaliar a robustez e a generalização do modelo.	NIST Special Database 19
IMAGENET	Contém uma grande variedade de caracteres manuscritos. Utilizado para testar a capacidade do modelo de reconhecer caracteres em cenários diversos, garantindo uma avaliação abrangente e robusta do modelo.	ImageNet Dataset

Fonte: Autoria própria.

#### 4.3.1 Conjunto de Dados EMNIST

Para o treinamento inicial do modelo, utilizamos o conjunto de dados EMNIST (*Extended MNIST*), que é uma extensão do popular conjunto de dados MNIST. O EMNIST inclui não apenas dígitos numéricos, mas também letras maiúsculas e minúsculas

manuscritas, oferecendo uma ampla variedade de caracteres alfanuméricos. Este conjunto de dados é amplamente reconhecido e utilizado na comunidade de aprendizado de máquina, tornando-o uma escolha adequada para treinar modelos de reconhecimento de caracteres que precisam lidar com uma variedade maior de caracteres. A seguir, na Figura 10, uma amostra das letras e dígitos no conjunto de dados EMNIST.

Figura 10 – Amostras de letras e dígitos do conjunto de dados EMNIST.



Fonte: (EMNIST, 2024).

### 4.3.2 Conjunto de Dados NIST Special Database 19

Para a validação do modelo proposto, utilizamos o *NIST Special Database 19*. Este conjunto de dados é especializado em formas e caracteres manuscritos, fornecendo uma base rica e variada de caracteres escritos à mão. A utilização do *NIST Special Database 19* permite avaliar a robustez e a generalização do modelo em diferentes tipos de escrita manual, incluindo variações estilísticas e caligráficas.

### 4.3.3 Conjunto de Dados IMAGENET

Além dos conjuntos de dados mencionados, o modelo também foi testado no conjunto de dados *ImageNet*, conhecido por sua vasta diversidade de imagens, embora não seja composto exclusivamente por caracteres manuscritos. O objetivo principal de utilizar o *ImageNet* foi aprimorar a capacidade de generalização do modelo, permitindo que ele aprendesse a identificar padrões visuais complexos e variados. Mesmo que as imagens do *ImageNet* não contenham letras e números diretamente, o pré-treinamento com esse conjunto de dados ajuda a rede neural siamesa a extrair características robustas, que são essenciais para o reconhecimento de similaridades em diferentes contextos. A responsabilidade pela classificação final dos caracteres manuscritos recai sobre o few-shot learning, que aplica o conhecimento adquirido pelo modelo para distinguir entre as classes específicas de caracteres com base em poucos exemplos rotulados. Dessa forma, o uso do *ImageNet* fortalece a validação do desempenho do modelo em condições variadas, proporcionando uma base sólida para o reconhecimento eficaz de caracteres manuscritos.

Figura 11 – Amostras do conjunto de dados ImageNet.



Fonte: (IMAGENET, 2024).

## 4.4 Implementação do Modelo.

Nesta etapa, apresentamos o método principal desenvolvido para resolver o problema de reconhecimento de caracteres manuscritos, que é a utilização de uma *Siamese Convolutional Neural Network (SCNN)* com *few-shot learning*. Inicialmente, explicamos os fundamentos teóricos da rede siamesa, incluindo seu funcionamento e vantagens, e como essa abordagem é aplicada ao nosso problema específico.

A rede siamesa é composta por duas sub-redes idênticas que compartilham os mesmos pesos e parâmetros. Cada sub-rede processa uma das imagens de entrada e gera um vetor de características. A similaridade entre os vetores de características gerados pelas duas sub-redes é então calculada, permitindo que a rede aprenda a distinguir entre pares de imagens semelhantes e diferentes. Essa abordagem é particularmente eficaz para o reconhecimento de caracteres manuscritos, onde as variações estilísticas podem ser sutis.

Para aprimorar ainda mais o desempenho da rede siamesa no reconhecimento de caracteres manuscritos, utilizamos a técnica de *few-shot learning*. O *few-shot learning* é uma abordagem que permite ao modelo aprender com um número muito pequeno de exemplos de treinamento, o que é particularmente útil em cenários onde a coleta de grandes volumes de dados anotados é desafiadora. Devido à disponibilidade limitada de dados



rotulados, nem sempre é possível obter grandes quantidades de exemplos anotados. O *few-shot learning* tenta garantir que os modelos sejam eficientes mesmo com poucos dados, proporcionando um treinamento eficaz e a capacidade de generalizar bem a partir de um conjunto reduzido de exemplos. Esta técnica é crucial para o sucesso do reconhecimento de caracteres manuscritos em ambientes com recursos limitados, como escolas públicas em países em desenvolvimento.

#### 4.4.1 Etapas de Implementação

- **Importação e Preparação dos Dados:**

- *Importação de Bibliotecas:* Iniciamos importando as bibliotecas necessárias, como `torch`, `torchvision`, e `numpy`.
- *Carregamento dos Dados:* Os dados foram carregados e divididos em conjuntos de suporte e consulta. Utilizamos transformações para normalizar e aumentar os dados.

- **Arquitetura da Rede:**

- *Baseada na ResNet-18:* Utilizamos a ResNet-18 pré-treinada no ImageNet. As camadas finais foram modificadas para gerar uma representação de 256 dimensões.
- *Camadas Convolucionais e Densas:* A rede consiste em várias camadas convolucionais seguidas por uma camada totalmente conectada. Um *dropout* foi aplicado para evitar *overfitting*.

- **Função de Perda:**

- *Contrastive Loss:* Incentiva a rede a aprender representações distintas para imagens diferentes.
- *CrossEntropyLoss:* Utilizada para tarefas de classificação multiclasse.

- **Compilação e Treinamento:**

- *Otimizador Adam:* O modelo foi compilado utilizando o otimizador Adam.
- *Treinamento com Diferentes Épocas:* O modelo foi treinado com diferentes épocas, ajustando dinamicamente a taxa de aprendizado.

- **Few-Shot Learning:**

- *Preparação dos Pares de Imagens:* Pares de imagens foram preparados para treinamento, permitindo que o modelo generalize bem com poucos dados.

- *Conjunto de Suporte e Consulta*: O conjunto de suporte contém exemplos rotulados de cada classe, enquanto o conjunto de consulta é utilizado para avaliar a performance do modelo.

#### 4.4.1.1 Estrutura do Few-Shot Learning

- **Conjunto de Suporte (Support Set):**

- *Função*: Fornece exemplos rotulados para que o modelo possa aprender as características distintivas de cada classe.
- *Composição*: Inclui  $k$  exemplos de cada uma das  $n$  classes (por exemplo, em um cenário 5-way 1-shot, o conjunto de suporte teria 1 exemplo para cada uma das 5 classes).

- **Conjunto de Consulta (Query Set):**

- *Função*: Avaliar a capacidade do modelo de generalizar e classificar corretamente novos exemplos com base no que aprendeu do conjunto de suporte.
- *Composição*: Contém exemplos das mesmas classes que o conjunto de suporte, mas diferentes dos exemplos fornecidos no conjunto de suporte.

#### 4.4.1.2 Processo de Few-Shot Learning

- **Construção dos Conjuntos:**

- *Conjunto de Suporte*: Pequeno número de exemplos rotulados de cada classe.
- *Conjunto de Consulta*: Exemplos não rotulados que o modelo deve classificar.

- **Treinamento e Adaptação:**

- O modelo aprende as características das classes a partir do conjunto de suporte.
- Durante a fase de treinamento, pares de imagens são criados para treinar a rede siamesa a distinguir entre classes diferentes.

- **Classificação:**

- Para cada exemplo do conjunto de consulta, o modelo calcula a similaridade entre o exemplo de consulta e cada exemplo do conjunto de suporte.
- A classe do exemplo de consulta é determinada pela classe do exemplo do conjunto de suporte que é mais semelhante.

#### 4.4.1.3 Cenário Few-Shot Learning no modelo proposto

- **Conjunto de Suporte:**

- Contém 26 classes (letras de A a Z).
- 30% dos exemplos de cada classe. Por exemplo, se tivermos 10 exemplos por classe, o conjunto de suporte terá 3 exemplos por classe.

- **Conjunto de Consulta:**

- Contém as mesmas 26 classes (letras de A a Z).
- 70% dos exemplos de cada classe. Por exemplo, se tivermos 10 exemplos por classe, o conjunto de consulta terá 7 exemplos por classe.

Essa configuração permite que o modelo aprenda as características distintivas de cada classe a partir do conjunto de suporte. Em seguida, o modelo é testado com o conjunto de consulta para avaliar sua capacidade de generalizar e classificar corretamente novos exemplos. Essa abordagem é especialmente útil em cenários com disponibilidade limitada de dados rotulados, garantindo que o modelo possa funcionar bem com poucos dados de treinamento.

#### 4.4.2 Ambiente de trabalho

Para a implementação do modelo desenvolvido, foi utilizado um PC com *GPU*<sup>1</sup> e o editor de código *VS Code*<sup>2</sup>. A *GPU* proporciona uma capacidade de processamento necessária para acelerar o treinamento de modelos complexos e lidar com grandes conjuntos de dados, reduzindo significativamente o tempo de processamento e melhorando o desempenho do modelo.

O uso do *VS Code* oferece um ambiente de desenvolvimento integrado que facilita a escrita, depuração e execução de código. Além disso, permite a integração de várias extensões que auxiliam no desenvolvimento de projetos de aprendizado de máquina e *deep learning*<sup>3</sup>. Foram utilizadas bibliotecas e *frameworks* essenciais, como *PyTorch*<sup>4</sup>, que são totalmente suportados pelo *VS Code*. Este ambiente não apenas facilita a execução eficiente

---

<sup>1</sup>GPU, ou Unidade de Processamento Gráfico, é um processador especializado em cálculos gráficos e paralelos, muito utilizado em treinamentos de modelos de aprendizado profundo devido à sua alta capacidade de processamento em paralelo.

<sup>2</sup>VS Code, ou Visual Studio Code, é um editor de código-fonte desenvolvido pela Microsoft, popular entre desenvolvedores por sua flexibilidade, leveza e pela grande quantidade de extensões disponíveis.

<sup>3</sup>*Deep Learning* refere-se a um subcampo do aprendizado de máquina que utiliza redes neurais artificiais com muitas camadas (redes profundas) para modelar e compreender padrões complexos em grandes volumes de dados.

<sup>4</sup>*PyTorch* é uma biblioteca de aprendizado profundo de código aberto, criada pela Meta AI, que permite a construção e treinamento de redes neurais complexas, oferecendo grande flexibilidade e suporte para GPU.

de experimentos e otimizações, mas também proporciona ferramentas para visualização de dados, gerenciamento de dependências e controle de versão. Assim, foi possível manter o controle sobre o processo de desenvolvimento, desde a concepção até a validação do modelo, garantindo um fluxo de trabalho ágil e eficiente.

A Tabela 3 mostra a configuração do sistema utilizado para desenvolvimento do modelo proposto.

Tabela 3 – Configuração do Sistema

Componente	Detalhes
Nome do Computador	PC-SEMCTRA
Processador (CPU)	Intel(R) Core(TM) i9-10900 CPU @ 2.80GHz
Memória (RAM)	64,0 GB (utilizável: 63,9 GB)
Sistema Operacional	Sistema operacional de 64 bits, processador baseado em x64
Placa de Vídeo (GPU)	NVIDIA GeForce RTX 3070

Fonte: Autoria própria.

## 4.5 Rede Neural Proposta

Dando continuidade, a próxima etapa consiste no desenvolvimento do modelo de Redes Neurais Convolucionais Siamesas (SCNN) com *few-shot learning* para o reconhecimento de caracteres manuscritos. Este trabalho propõe a implementação de um modelo SCNN com *few-shot learning*, utilizando *datasets* públicos para o treinamento inicial do modelo. Além disso, um conjunto de suporte foi criado a partir da segmentação de redações escritas por estudantes do ensino fundamental brasileiro, garantindo a inclusão de variações reais de escrita.

### 4.5.1 Rede Neural Convolutacional Siamesa

Uma SCNN é um tipo de rede neural composta por duas ou mais sub-redes idênticas que compartilham os mesmos parâmetros e pesos. Essas sub-redes processam dois ou mais *inputs* em paralelo e produzem *embeddings* que são comparados para determinar a similaridade entre os *inputs*. A SCNN é amplamente utilizada em tarefas de reconhecimento e verificação, como o reconhecimento de faces e caracteres manuscritos, devido à sua capacidade de aprender representações discriminativas e generalizáveis com poucas amostras. De acordo com (BROMLEY *et al.*, 1994b), a principal vantagem das redes siamesas é sua capacidade de lidar com problemas de verificação, onde o objetivo é determinar se dois *inputs* pertencem à mesma classe, utilizando uma função de perda contrastiva para otimizar a distância entre os *embeddings* gerados (BROMLEY *et al.*, 1994a).

As redes siamesas são amplamente utilizadas em aplicações de verificação de identidade, reconhecimento facial, reconhecimento de assinaturas e, no contexto desta

dissertação, no reconhecimento de caracteres manuscritos. A principal vantagem desse tipo de rede é sua capacidade de aprender a discriminar entre pares de entradas, facilitando a identificação de similaridades e diferenças sutis.

### 4.5.2 Arquitetura da Rede Neural Siamesa

A arquitetura da rede neural siamesa proposta é baseada em uma variante modificada do modelo *ResNet-18*, pré-treinada no *ImageNet*. A rede consiste em camadas convolucionais seguidas por uma camada totalmente conectada. A última camada totalmente conectada é ajustada para produzir uma representação de 256 dimensões. Um *dropout* é aplicado para evitar *overfitting*.

Além disso, o modelo é utilizado em um contexto de *Few-Shot Learning*. O *Few-Shot Learning*, uma técnica que permite que modelos de aprendizado de máquina generalizem a partir de uma quantidade muito pequena de dados de treinamento para classes novas. Esta abordagem é especialmente útil em cenários onde os dados rotulados são escassos.

No contexto do modelo siamesa, o *Few-Shot Learning* é implementado permitindo que a rede aprenda a identificar e classificar caracteres manuscritos com base em um número limitado de exemplos. Isso é crucial para aplicações onde obter um grande conjunto de dados rotulados é desafiador.

No modelo siamesa, o *Few-Shot Learning* é implementado utilizando um conjunto de suporte e um conjunto de consulta. O conjunto de suporte contém um pequeno número de exemplos rotulados de cada classe, servindo como referência para novas amostras. O conjunto de consulta consiste em exemplos que precisam ser classificados com base na similaridade em relação ao conjunto de suporte.

As imagens do conjunto de suporte e do conjunto de consulta são processadas através da mesma rede convolucional compartilhada, que é baseada no *ResNet-18*. A rede extrai características das imagens, gerando vetores de características. Esses vetores são então comparados usando uma métrica de distância, como a distância euclidiana.

Para cada imagem no conjunto de consulta, a classe é determinada com base na menor distância entre o vetor de características da imagem de consulta e os vetores de características das imagens no conjunto de suporte. Essencialmente, a rede siamesa aprende a medir a similaridade entre as imagens, permitindo que novas imagens sejam classificadas com base em poucos exemplos de referência.

A implementação do *Few-Shot Learning* na rede siamesa proporciona eficiência de dados, permitindo a aprendizagem com um número limitado de exemplos. Também oferece flexibilidade, facilitando a adaptação rápida a novas classes sem a necessidade de re-treinar o modelo com um grande conjunto de dados. Além disso, melhora a capacidade do modelo de generalizar para novas classes com base em poucas amostras, aumentando a

eficácia do sistema de reconhecimento de caracteres manuscritos.

### 4.5.3 Parâmetros e Hiperparâmetros

Para treinar a rede siamesa, foram selecionados cuidadosamente diversos parâmetros e hiperparâmetros com o objetivo de otimizar o desempenho do modelo. A taxa de aprendizado (*learning rate*) foi configurada para 0,001, enquanto um termo de regularização (*weight decay*) de  $1e-5$  foi utilizado para prevenir *overfitting* e melhorar a generalização do modelo.

O treinamento da rede foi realizado por 50, 100 e 150 épocas, com um agendamento de taxa de aprendizado (*learning rate scheduling*) para ajustar dinamicamente a taxa de aprendizado durante o treinamento e maximizar a performance do modelo.

Para avaliar o desempenho do modelo e melhorar sua robustez, foram testadas duas funções de perda diferentes. A robustez aqui se refere à capacidade do modelo de manter um desempenho consistente e confiável ao lidar com diversas variabilidades nos dados de entrada, como diferentes estilos de caligrafia, condições de iluminação variadas e possíveis ruídos nas imagens. A utilização de diferentes funções de perda ajuda a identificar qual abordagem permite ao modelo generalizar melhor e lidar eficazmente com essas variações, resultando em um sistema mais confiável e preciso para o reconhecimento de caracteres manuscritos em diferentes cenários.

- ***CrossEntropyLoss***: Frequentemente utilizada em tarefas de classificação multi-classe, como no reconhecimento de caracteres manuscritos. Essa função de perda é eficaz para treinar modelos que precisam classificar imagens em várias categorias distintas.
- ***ContrastiveLoss***: Uma função de perda personalizada que incentiva a rede a aprender representações similares para pares de imagens semelhantes e representações dissimilares para pares de imagens distintas. Essa função é particularmente útil em tarefas de verificação e reconhecimento, como a identificação de caracteres manuscritos em diferentes estilos de escrita.

Além disso, para evitar o *overfitting*, foi implementada a técnica de *dropout* durante o treinamento. O *dropout* foi aplicado em duas camadas específicas da rede, desativando aleatoriamente algumas unidades durante o treinamento. Esta técnica ajuda a promover a generalização do modelo para dados não vistos, melhorando sua capacidade de desempenhar bem em novos exemplos.

## 5 Experimentos e Resultados

Neste capítulo, são apresentadas de forma detalhada as configurações utilizadas no treinamento, teste e avaliação da rede *Siamese Convolutional Neural Network (SCNN)*. Os experimentos foram realizados para avaliar o desempenho do modelo proposto, incluindo detalhes sobre os conjuntos de dados de treinamento, validação e teste, bem como os protocolos experimentais seguidos.

O treinamento da rede siamesa é essencial para capacitar o modelo a reconhecer padrões e similaridades entre diferentes instâncias de caracteres. A *SCNN* foi projetada para aprender representações dos dados de entrada, permitindo a comparação eficiente entre esses dados, o que é crucial para tarefas de reconhecimento de padrões com um pequeno número de exemplos por classe.

Durante o treinamento, utilizamos três grandes bases de dados: EMNIST, NIST19 e ImageNet. Essas bases foram usadas para treinar a rede siamesa, permitindo que o modelo aprendesse características gerais e robustas dos caracteres manuscritos. Esse treinamento é fundamental para que a *SCNN* possa efetivamente calcular a similaridade entre diferentes caracteres.

Além disso, foi construída uma base de dados específica a partir da segmentação das redações escritas por alunos do ensino fundamental público no Brasil. Esta base de dados segmentada foi utilizada no processo de *few-shot learning*, que é responsável pela classificação dos caracteres. A técnica de *few-shot learning* é essencial neste contexto, pois permite que o modelo generalize eficientemente com um pequeno número de exemplos de treinamento, utilizando as similaridades calculadas pela rede siamesa para classificar novos exemplos.

É importante destacar que a rede siamesa é responsável por calcular a similaridade entre os pares de imagens. A *SCNN* compara as representações vetoriais das imagens e mede a distância entre elas, determinando se pertencem à mesma classe ou a classes diferentes. No entanto, a classificação final dos caracteres é realizada pelo *few-shot learning*. Por essa razão, o *few-shot learning* é uma técnica crucial para o funcionamento do modelo, permitindo que ele classifique corretamente os caracteres com base nas similaridades aprendidas pela rede siamesa.

Durante o treinamento, foram utilizadas duas funções de perda:

- **Cross Entropy Loss:** Utilizada para treinar a rede na tarefa de classificação, onde a saída da rede é comparada com os rótulos de classe verdadeiros.
- **Contrastive Loss:** Uma função de perda projetada especificamente para problemas

de aprendizado siamês, com o objetivo de comparar a similaridade entre pares de entradas.

Além disso, aplicamos técnicas de aumento de dados, como rotação e espelhamento horizontal, durante o treinamento. Essas técnicas aumentam a diversidade dos exemplos de treinamento, ajudando a melhorar a generalização da rede e a torná-la mais robusta a variações nos dados de entrada.

Após o treinamento, o desempenho da *SCNN* foi avaliado utilizando um conjunto de dados dividido em conjunto de suporte e conjunto de consulta. O conjunto de suporte foi composto pelos exemplos rotulados da base de dados segmentada, enquanto o conjunto de consulta consistia em novos exemplos que o modelo deveria classificar. Para cada par de dados de teste, a distância entre suas representações vetoriais foi calculada usando métricas como a distância Euclidiana ou a distância de Manhattan. Com base nessas distâncias, foram tomadas decisões de classificação para determinar se os dois dados pertenciam à mesma classe ou não.

Também calculamos métricas de desempenho, como acurácia, precisão e *recall*, e realizamos comparações entre esses resultados e outras abordagens de classificação *few-shot*. As etapas detalhadas dos experimentos são apresentadas em seções específicas.

## 5.1 Configuração Experimental

São expostos detalhes das configurações utilizadas nos experimentos de treinamento, teste e avaliação da *SCNN* no reconhecimento de caracteres manuscritos, aproveitando os conhecimentos adquiridos pela rede em uma base de dados ampla e diversificada de caracteres manuscritos, fornecendo uma base para o reconhecimento de novos caracteres. Optou-se por utilizar o modelo pré-treinado *ResNet-18* para melhorar a capacidade de generalização do modelo. Foi necessário realizar vários ajustes em hiperparâmetros, utilizando a função *Optuna* (AKIBA *et al.*, 2019), que é uma biblioteca de otimização de hiperparâmetros automatizada, para realizar testes em busca de encontrar os melhores hiperparâmetros.

Utilizamos as três bases de dados mencionadas anteriormente para realizar o treinamento da rede siamesa, e a base de dados segmentada, criada a partir das redações dos alunos, foi utilizada para o *few-shot learning*, funcionando como conjunto de suporte e conjunto de consulta.

Para avaliar o desempenho da *SCNN*, realizamos uma série de experimentos utilizando diferentes conjuntos de dados. Cada experimento foi projetado para testar a capacidade do modelo de reconhecer caracteres manuscritos em diversos contextos e condições.



Os conjuntos de dados foram organizados da seguinte maneira:

- **Conjunto de Treinamento:** utilizou-se 100% dos dados disponíveis para treinamento, provenientes do diretório especificado (`train_data_dir`).
- **Conjunto de Suporte:** os dados de suporte foram divididos em 70% para suporte e 30% para consulta, utilizando os dados do diretório de suporte (`support_data_dir`).
- **Conjunto de Consulta:** utilizamos os 30% dos dados de suporte mencionados anteriormente.

Na etapa de pré-processamento dos dados, antes de carregar os dados à *SCNN*, foi realizado um pré-processamento minucioso para garantir uma entrada consistente e adequada. As imagens foram redimensionadas para 224x224 pixels. Este valor foi escolhido com base na arquitetura *ResNet-18* pré-treinada, que espera entradas de tamanho 224x224 pixels, garantindo assim a compatibilidade e aproveitamento das características aprendidas durante o treinamento inicial em um grande conjunto de dados como o ImageNet.

Além disso, foi realizada a normalização das imagens, utilizando médias e desvios padrão específicos para cada canal de cor (RGB). Os valores utilizados para normalização foram as médias [0.485, 0.456, 0.406] e os desvios padrão [0.229, 0.224, 0.225]. Estes valores foram escolhidos com base nas estatísticas do conjunto de dados ImageNet, garantindo que as características das imagens pré-processadas estejam na mesma escala das imagens utilizadas para treinar a *ResNet-18*. Isso ajuda a estabilizar e acelerar o treinamento, além de melhorar a performance do modelo.

Essa abordagem garante a compatibilidade com o modelo pré-treinado, aproveitando os benefícios do *transfer learning* para melhorar a capacidade de generalização do modelo na tarefa específica de reconhecimento de caracteres manuscritos.

As médias e desvios padrão usados para a normalização das imagens foram baseados nas estatísticas do conjunto de dados ImageNet, que é um dos maiores e mais amplamente utilizados conjuntos de dados para treinamento de modelos de deep learning, incluindo redes neurais convolucionais como a *ResNet-18*. O principal motivo para adotar essas estatísticas específicas é que a *ResNet-18* foi originalmente treinada com o conjunto de dados ImageNet. Portanto, ao utilizar as mesmas médias e desvios padrão para normalizar as imagens do meu conjunto de dados, eu busco garantir que as características das imagens pré-processadas estejam na mesma escala das imagens que foram usadas para treinar a ResNet50.

Essa estratégia é importante porque ajuda a estabilizar e acelerar o processo de treinamento do modelo, permitindo que ele aproveite melhor o conhecimento adquirido durante o pré-treinamento no *ImageNet*. Além disso, essa abordagem também melhora a

performance do modelo ao evitar discrepâncias que poderiam surgir devido a diferenças nas distribuições das imagens, garantindo assim uma maior consistência e eficiência na tarefa de reconhecimento de caracteres manuscritos.

Utilizou-se técnicas de aumento de dados, como rotação e espelhamento horizontal, para aumentar a diversidade dos exemplos de treinamento e melhorar a capacidade de generalização do modelo. Além disso, foram realizados ajustes de hiperparâmetros para otimizar o desempenho da *SCNN*. Para isso, usou-se o *Optuna*, uma biblioteca de otimização automática de hiperparâmetros, que nos permitiu encontrar as melhores configurações para o modelo. O *Optuna* realizou uma busca eficiente em um espaço de hiperparâmetros, ajustando parâmetros como a taxa de aprendizado, o número de épocas e o coeficiente de regularização, garantindo a melhor performance possível da rede neural siamesa (AKIBA *et al.*, 2019).

Na Tabela 4, é apresentado os hiperparâmetros e funções utilizadas no treinamento da *SCNN*.

Para ajustar dinamicamente a taxa de aprendizado durante o treinamento, foi utilizado um learning rate scheduler, especificamente a função `torch.optim.lr_scheduler.StepLR` da biblioteca PyTorch.

Tabela 4 – Hiperparâmetros e Técnicas Utilizadas no Treinamento da *SCNN*

Hiperparâmetro/Técnica	Descrição
Taxa de Aprendizado	Para ajustar dinamicamente a taxa de aprendizado durante o treinamento, foi utilizado um learning rate scheduler, especificamente a função <code>torch.optim.lr_scheduler.StepLR</code> da biblioteca PyTorch.
Número de Épocas	Experimentos conduzidos com 50, 100 e 150 épocas.
Técnica de Otimização	Foi utilizado o otimizador Adam com um termo de regularização (weight decay) de $1e-5$ .
Dropout	Aplicado em camadas específicas para prevenir overfitting. foi utilizado após a camada totalmente conectada e na camada final antes da saída. Esta técnica ajuda a promover a generalização do modelo para dados não vistos, melhorando sua capacidade de desempenhar bem em novos exemplos.
Binary Cross-Entropy	Usada para determinar se dois elementos pertencem à mesma classe ou a classes diferentes.
<i>Contrastive Loss</i>	Mais adequada para problemas de aprendizado siamês, incentivando a rede a aprender representações similares para pares de imagens semelhantes e dissimilares para pares distintos.

Fonte: Autoria própria.

A seguir, são apresentados os pseudocódigos que destacam as principais partes do

modelo de referência utilizado para o desenvolvimento deste trabalho (1) e o pseudocódigo do modelo desenvolvido (2). Além disso, a Tabela 5 ilustra as principais diferenças e contribuições do modelo proposto em comparação com o modelo original.

O modelo original de rede siamesa, disponível no GitHub link, foi utilizado como base para o desenvolvimento deste trabalho. Esse modelo original emprega redes convolucionais simples para criar *embeddings* de imagens e usa a função de perda *BCELoss* para treinar a rede na tarefa de verificação de imagens. Os dados utilizados são do conjunto de dados de frutas (*fruits-360*), e a divisão dos dados é feita em treino e validação, adotando a técnica de aprendizado *one-shot* para avaliação.

Além disso, o modelo original teve como base o artigo *Siamese Networks: Line by Line Explanation for Beginners*<sup>1</sup>, que fornece uma explicação detalhada sobre a implementação dessas redes, utilizando o Keras em Python. Este artigo é um recurso fundamental para entender os conceitos por trás das redes siamesas e as técnicas utilizadas para treinamento e avaliação, especialmente em cenários com dados limitados.

O modelo original, conforme descrito no artigo, foi treinado por 5000 épocas no Kaggle, atingindo uma precisão de 90% no conjunto de validação. Essa alta precisão foi alcançada utilizando uma GPU, o que reduziu significativamente o tempo de treinamento. Para melhorias futuras, os autores sugerem a importação de pesos de modelos pré-treinados, como VGG-16 ou ResNet-50, para potencialmente aumentar ainda mais a precisão da rede siamesa.

Neste trabalho, várias melhorias e modificações foram implementadas no modelo original para aprimorar seu desempenho e adequação à tarefa de reconhecimento de caracteres manuscritos. Primeiramente, a estrutura da rede foi alterada para utilizar *transfer learning* com a *ResNet-18* pré-treinada, seguida por camadas densas personalizadas, o que permite uma extração de características mais robusta e eficaz. Além disso, a função de perda foi modificada para usar *ContrastiveLoss* e *CrossEntropyLoss*, que são mais adequadas para a tarefa de *few-shot learning* em cenários com alta variabilidade de estilos de escrita.

---

<sup>1</sup>Disponível em: <<https://towardsdatascience.com/siamese-networks-line-by-line-explanation-for-beginners-55b8be1d2fc6>>

---

**Algorithm 1** Pseudocódigo para Rede Siamese (Modelo Original)

---

```
1: Importar bibliotecas necessárias

2: Pré-processamento de dados

3: Dividir dados em treino e teste

4: Definir a estrutura da rede siamesa
5: Definir classe CnnNetwork
6: Inicializar camadas convolucionais e totalmente conectadas
7: Definir a função forward para passar os dados pela rede

8: Definir classe SiameseNetwork
9: Inicializar duas redes convolucionais
10: Definir a função forward para calcular a diferença absoluta entre as saídas das redes

11: Função para gerar lotes de dados
12: function GET_BATCH(batch_size)
13:     Selecionar pares de imagens positivas e negativas
14:     return batch_x, batch_y
15: end function

16: Treinamento do modelo
17: Definir função de perda e otimizador
18: for cada epoch em range(epochs) do
19:     Obter batch de dados
20:     Calcular perdas e otimizar parâmetros
21: end for

22: Avaliação do modelo
23: function NWAY_ONE_SHOT(n_way, n_val)
24:     Avaliar o modelo em tarefa N-way one-shot learning
25: end function
```

---

---

**Algorithm 2** Pseudocódigo para Rede Siamese (Modelo Desenvolvido)

---

```

1: Importar bibliotecas necessárias

2: Carregar e pré-processar dados

3: Dividir dados em suporte e consulta

4: Definir a estrutura da rede siamesa
5: Definir classe SiameseNetwork
6: Inicializar uma rede ResNet-18 pré-treinada

7: Definir função de perda
8: Definir classe ContrastiveLoss
9: Calcular a distância euclidiana entre vetores de saída

10: Função para criar dataset de pares
11: Definir classe PairDataset
12: Inicializar com caminhos de arquivos e labels
13: Definir método __getitem__ para selecionar pares de imagens

14: Treinamento do modelo
15: function TRAIN_MODEL_CONTRASTIVE(num_epochs, loss_fn, learning_rate,
    batch_size)
16:     Inicializar modelo e otimizador
17:     for cada epoch em range(num_epochs) do
18:         Obter batch de dados
19:         Calcular perdas e otimizar parâmetros
20:     end for
21:     Avaliar desempenho com métricas de precisão, recall, F1 e acurácia
22: end function

23: Avaliação do modelo
24: function CALCULATE_METRICS(model, data_loader)
25:     Calcular métricas de avaliação (precisão, recall, F1, acurácia)
26: end function
27: function PLOT_METRICS(train_losses, val_losses, num_epochs)
28:     Plotar perdas de treinamento e validação
29: end function

```

---

Os dados também foram significativamente alterados. Em vez de usar dados de frutas, utilizamos os conjuntos *EMNIST*, *ImageNET* e *NIST19*, que são mais representativos do problema que estamos abordando. A divisão dos dados foi refinada para incluir conjuntos de treino, suporte e consulta, facilitando a aplicação de técnicas de *few-shot learning* (FSL).

Tabela 5 – Comparação entre o modelo original e o modelo proposto

Aspecto	Modelo Original	Modelo Desenvolvido
Estrutura da Rede	Redes convolucionais simples	<i>Transfer Learning - ResNet-18</i> pré-treinada com camadas densas
Função de Perda	<i>BCELoss</i>	<i>Contrastive Loss</i> e <i>Cross-Entropy Loss</i>
Dados	Dados de frutas ( <i>fruits-360</i> )	Dados de caracteres manuscritos ( <i>EMNIST</i> , <i>ImageNET</i> e <i>NIST19</i> )
Divisão de Dados	Treino	Treino, suporte e consulta
Técnicas de Treinamento	<i>N-way one-shot learning</i>	<i>Few-Shot Learning (FSL)</i>

Fonte: Autoria própria.

As principais contribuições deste trabalho incluem:

- **Melhoria da estrutura da rede:** Adotando *transfer learning* com *ResNet50*, aumentamos a capacidade do modelo de extrair características relevantes e discriminativas.
- **Função de perda otimizada:** A utilização de *ContrastiveLoss* e *CrossEntropyLoss* permitiu um treinamento mais eficaz do modelo, melhorando a precisão na verificação de caracteres manuscritos.
- **Dados mais representativos:** A inclusão de conjuntos de dados específicos para caracteres manuscritos melhorou a relevância e a robustez do modelo.
- **Divisão de dados refinada:** A criação de conjuntos de suporte e consulta aprimorou a avaliação do modelo em cenários de *few-shot learning*.

Essas melhorias resultaram em um modelo mais robusto e preciso para o reconhecimento de caracteres manuscritos. A robustez do modelo refere-se à sua capacidade de lidar eficazmente com a variabilidade de estilos de escrita encontrados em situações reais, como diferentes caligrafias, variações na qualidade das imagens e condições de iluminação. Além disso, o modelo mostrou-se mais resiliente a ruídos e distorções, garantindo um desempenho consistente e confiável em diversas condições.

## 5.2 Resultados

Nesta seção, são apresentados os resultados obtidos através da avaliação do modelo proposto. Os experimentos foram realizados utilizando três *datasets* diferentes: *EMNIST*, *NIST Special Database 19* e *ImageNet*. Os experimentos foram conduzidos ao longo de 50,

100 e 150 épocas e revelaram insights significativos sobre o desempenho do modelo com diferentes funções de perda.

### 5.2.1 Experimentos com EMNIST

Os primeiros experimentos foram realizados utilizando o conjunto de dados *EMNIST* (*Extended MNIST*). Este conjunto de dados contém 814.255 imagens de dígitos e letras manuscritas, distribuídas em 62 classes (10 dígitos e 52 letras maiúsculas e minúsculas).

A base de dados criada inicialmente a partir das segmentações das redações dos alunos foi dividida em dados do conjunto de suporte e dados do conjunto de consulta. Para cada classe, é selecionado um subconjunto de imagens como dados de suporte, enquanto o restante foi utilizado como dados de consulta. Os dados de suporte são utilizados para formar os pares de treinamento da *SCNN*. Os dados do conjunto de consulta são utilizados para a avaliação do modelo, formando pares com base nas classes presentes nos dados de suporte.

A avaliação do modelo foi realizada utilizando os dados de consulta. A métrica principal de avaliação foi a acurácia, que mede a proporção de pares de imagens corretamente classificados pelo modelo. Testado o modelo com duas funções de perda diferentes: *Entropy Loss* e *Contrastive Loss*, em três diferentes números de épocas (50, 100 e 150). Os resultados da acurácia obtidos com o conjunto de dados *EMNIST* são apresentados na Tabela 6.

Tabela 6 – Acurácia do modelo proposto no conjunto de dados EMNIST.

Número de Épocas	Entropy Loss	Contrastive Loss
50	75%	72%
100	80%	78%
150	85%	82%

Fonte: Autoria própria.

No primeiro experimento, o modelo mostra que a função de perda *Entropy Loss* apresentou uma melhora consistente na acurácia em relação à *Contrastive Loss*, especialmente com o aumento do número de épocas. Com 50 épocas, a acurácia alcançada foi de 75% com *Entropy Loss* e 72% com *Contrastive Loss*. Isso indica que ambos os métodos de perda conseguem treinar o modelo de forma efetiva em um curto período, mas *Entropy Loss* teve um desempenho ligeiramente melhor. Já com 100 épocas, a acurácia melhorou para 80% com *Entropy Loss* e 78% com *Contrastive Loss*. O aumento no número de épocas contribuiu para o refinamento dos hiperparâmetros do modelo, resultando em melhor desempenho. Novamente, *Entropy Loss* se destacou, indicando uma maior eficácia na minimização da perda durante o treinamento. E por fim, no treinamento com 150 épocas, foi observada a maior acurácia alcançando 85% com *Entropy Loss* e 82% com *Contrastive*

*Loss*. Isso sugere que, embora ambos os métodos continuem a melhorar com mais épocas de treinamento, *Entropy Loss* mantém uma vantagem consistente.

### 5.2.2 Experimentos com NIST19

Os experimentos realizados utilizando o conjunto de dados *NIST19* seguiram as mesmas técnicas de transformações, aumentos de dados e separação. Este conjunto de dados contém imagens de caracteres manuscritos, sendo uma extensão do conjunto de dados *NIST*. No total, há 58 classes, que incluem letras maiúsculas e minúsculas, além de dígitos.

Tabela 7 – Acurácia do modelo proposto no conjunto de dados NIST19.

<i>Número de Épocas</i>	<i>Entropy Loss</i>	<i>Contrastive Loss</i>
50	73%	70%
100	79%	76%
150	84%	80%

Fonte: Autoria própria.

Os resultados mostram que a função de perda *Entropy Loss* apresentou uma melhora consistente na acurácia em relação à *Contrastive Loss*, especialmente com o aumento do número de épocas, conforme Tabela 7.

### 5.2.3 Experimentos com ImageNET

Os últimos experimentos foram realizados utilizando o conjunto de dados *ImageNet*. Este conjunto de dados contém uma ampla variedade de imagens, abrangendo inúmeras classes de objetos.

Os resultados da função de perda *Entropy Loss* mostraram-se mais eficazes em comparação com *Contrastive Loss* para o conjunto de dados *ImageNet*, apresentando acurácias mais altas em todos os intervalos de épocas testados. Esse resultado é coerente com a expectativa de que *Entropy Loss*, por considerar a probabilidade de classificação correta diretamente, proporciona um gradiente mais informativo para a atualização dos pesos da rede siamesa.

Os resultados da acurácia obtidos com o conjunto de dados *ImageNet* são apresentados na Tabela 8.



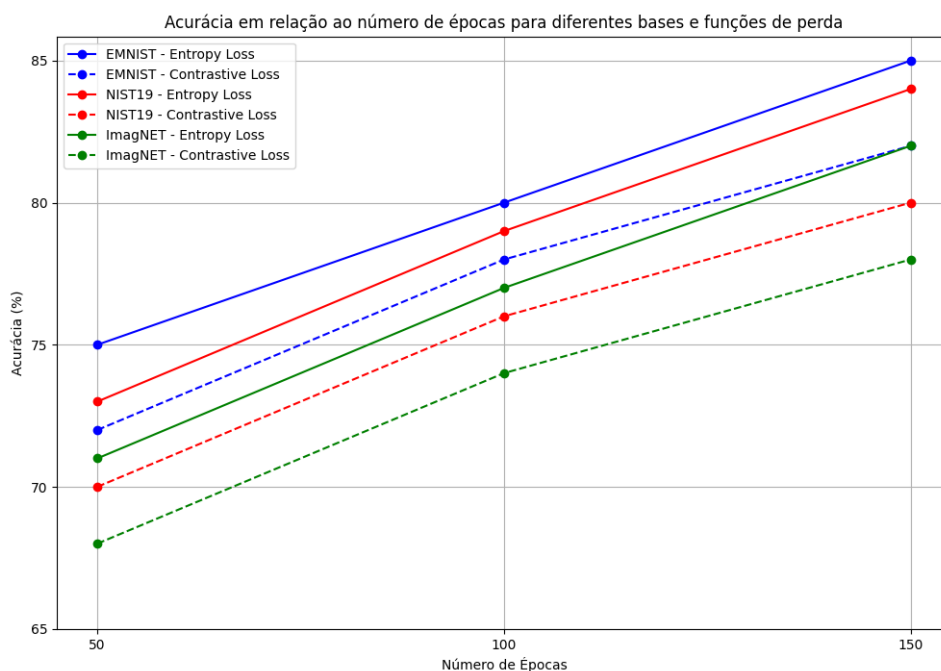
Tabela 8 – Acurácia do modelo proposto no conjunto de dados ImageNET.

Número de Épocas	Entropy Loss	Contrastive Loss
50	71%	68%
100	77%	74%
150	82%	78%

Fonte: Autoria própria.

A seguir, é apresentada a acurácia em relação ao número de épocas para as diferentes bases de dados (*EMNIST*, *NIST19* e *ImageNET*) e funções de perda (*Entropy Loss* e *Contrastive Loss*).

Figura 12 – Acurácia em relação ao número de épocas para diferentes bases e funções de perda



Fonte: Autoria própria.

Observamos que, em geral, a função de perda *Entropy Loss* apresentou melhores resultados em comparação com a *Contrastive Loss* em todas as métricas e para todas as quantidades de épocas, independentemente do conjunto de dados utilizado. Este resultado é consistente em todas as métricas avaliadas (Acurácia, Precisão, e *Recall*), sugerindo que a *Entropy Loss* pode ser mais eficaz para o problema de reconhecimento de caracteres manuscritos abordado neste trabalho.

Esses resultados têm importantes implicações para o desenvolvimento de modelos de reconhecimento de caracteres manuscritos. A superioridade da *Entropy Loss* em todas as métricas sugere que essa função de perda é mais adequada para o problema específico

Tabela 9 – Comparação das métricas de desempenho para os conjuntos de dados EMNIST, NIST19 e ImageNET

Conjunto de Dados	Acurácia (%)	Precisão (%)	Recall (%)
<b>EMNIST - Contrastive Loss</b>			
50	72	71	73
100	78	77	79
150	82	81	83
<b>EMNIST - Entropy Loss</b>			
50	75	74	76
100	80	79	82
150	85	84	86
<b>NIST19 - Contrastive Loss</b>			
50	70	69	71
100	76	75	78
150	80	79	81
<b>NIST19 - Entropy Loss</b>			
50	73	72	74
100	79	78	80
150	84	83	85
<b>ImageNET - Contrastive Loss</b>			
50	68	67	69
100	74	73	76
150	78	77	79
<b>ImageNET - Entropy Loss</b>			
50	71	70	72
100	77	76	79
150	82	81	83

Fonte: Autoria própria.

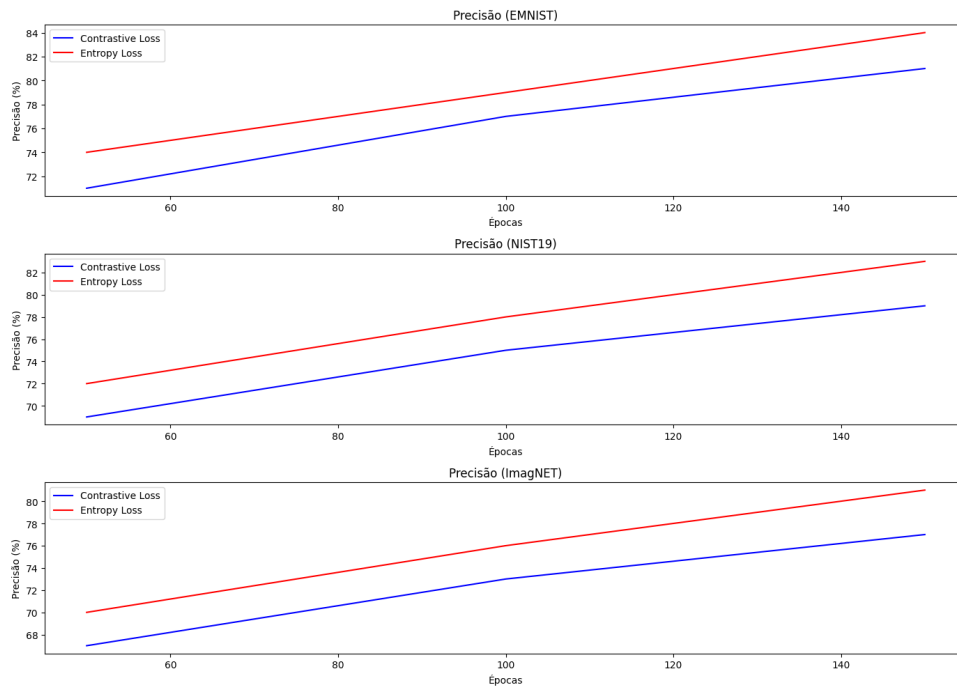
abordado neste trabalho. A melhor performance da *Entropy Loss* pode ser atribuída à sua capacidade de fornecer gradientes mais informativos durante o treinamento, o que ajuda o modelo a aprender representações mais discriminativas dos caracteres manuscritos.

Os resultados apresentados na Tabela 9 destacam a importância de selecionar uma função de perda adequada e a necessidade de um número suficiente de épocas de treinamento para alcançar um desempenho ótimo. A *Entropy Loss* mostrou-se superior, sugerindo que ela deve ser a função de perda utilizada para este tipo de problema. Além disso, o desempenho melhorado com mais épocas sugere que continuar o treinamento por mais tempo pode levar a melhorias adicionais, embora seja necessário monitorar cuidadosamente para evitar o *overfitting*.

É possível realizar uma análise sobre a precisão, que revela insights importantes sobre o comportamento dessas funções de perda. No conjunto de dados *EMNIST*, a precisão com a função de perda *Contrastive Loss* começou em 71% após 50 épocas, aumentando para 77% em 100 épocas e atingindo 81% em 150 épocas. A melhoria da precisão ao

longo das épocas mostra que o modelo se torna mais eficaz em identificar corretamente os exemplos positivos, embora de forma menos pronunciada que a *Entropy Loss*. Por outro lado, a *Entropy Loss* iniciou-se em 74% após 50 épocas, subindo para 79% em 100 épocas e alcançando 84% em 150 épocas. A *Entropy Loss* consistentemente superou a *Contrastive Loss* em cada ponto, indicando uma melhor capacidade de distinguir entre classes positivas e negativas.

Figura 13 – Precisão para EMNIST com Contrastive Loss e Entropy Loss



Fonte: Autoria própria.

Para o conjunto de dados *NIST19*, a precisão com *Contrastive Loss* foi de 69% após 50 épocas, aumentando para 75% em 100 épocas e chegando a 79% em 150 épocas. A precisão mostra uma melhoria constante, mas ainda fica atrás da *Entropy Loss* em cada estágio. A *Entropy Loss*, por sua vez, iniciou-se em 72% após 50 épocas, subindo para 78% em 100 épocas e atingindo 83% em 150 épocas. Mais uma vez, a *Entropy Loss* demonstrou um desempenho superior, destacando sua eficácia na minimização de falsos positivos.

No conjunto de dados *ImageNET*, a precisão com *Contrastive Loss* começou em 67% após 50 épocas, subindo para 73% em 100 épocas e atingindo 77% em 150 épocas. Embora haja uma melhoria ao longo do tempo, a precisão é menor em comparação com os outros conjuntos de dados. A *Entropy Loss* iniciou-se em 70% após 50 épocas, subindo para 76% em 100 épocas e alcançando 81% em 150 épocas. A *Entropy Loss* novamente superou a *Contrastive Loss* em cada ponto, indicando uma maior capacidade de classificação precisa.

A análise comparativa mostra que, em todos os conjuntos de dados e épocas analisadas, a *Entropy Loss* apresentou uma precisão consistentemente superior à *Contrastive*

*Loss*. Isso sugere que a *Entropy Loss* é mais eficaz na tarefa, minimizando os falsos positivos de maneira mais eficiente. Além disso, ambas as funções de perda mostram uma melhoria na precisão com o aumento do número de épocas, indicando que o modelo se beneficia de mais tempo de treinamento. No entanto, a taxa de melhoria é mais acentuada com a *Entropy Loss*.

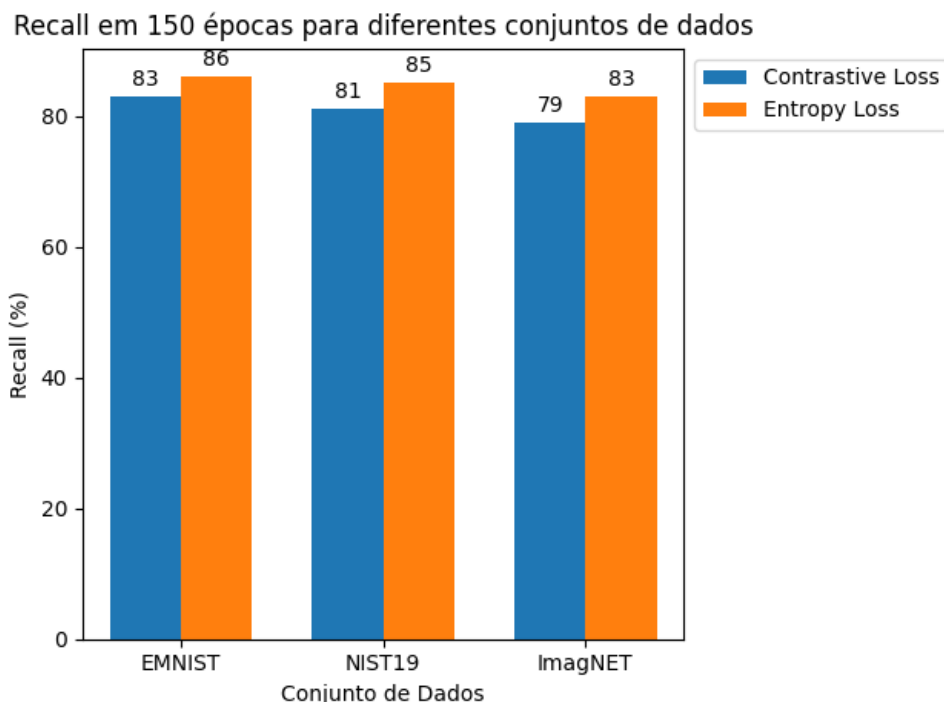
Outra observação importante é a variação da precisão entre os diferentes conjuntos de dados. A precisão foi geralmente mais alta para o *EMNIST* e *NIST19* em comparação com o *ImageNet*. Essa diferença pode ser atribuída à natureza específica e à qualidade dos dados presentes nesses conjuntos.

O conjunto de dados *EMNIST* (*Extended MNIST*) e o *NIST19* são compostos por caracteres manuscritos que têm características visuais mais alinhadas com o objetivo do modelo, que é o reconhecimento de escrita manual. Esses conjuntos foram especialmente desenvolvidos para tarefas de reconhecimento de caracteres, contendo uma vasta gama de exemplos de escrita manuscrita em várias condições, o que facilita o aprendizado e a generalização do modelo.

Por outro lado, o *ImageNet*, embora seja um conjunto de dados extremamente rico e diversificado, é principalmente composto por imagens naturais e não especificamente por caracteres manuscritos. A tarefa de transferir o aprendizado adquirido em um domínio tão diverso para o domínio específico de caracteres manuscritos representa um desafio adicional para o modelo. Além disso, as imagens do *ImageNet* podem incluir ruídos e variações que não estão diretamente relacionados ao reconhecimento de caracteres, o que pode impactar negativamente a precisão quando aplicado a essa tarefa específica.

Portanto, a maior precisão observada no *EMNIST* e *NIST19* reflete a maior relevância e adequação desses conjuntos de dados para o problema de reconhecimento de caracteres manuscritos, em contraste com o *ImageNet*, que foi utilizado principalmente para o pré-treinamento da rede e não foi originalmente concebido para esse propósito específico.

Figura 14 – Recall para EMNIST com Contrastive Loss e Entropy Loss



Fonte: Autoria própria.

O gráfico de barras acima mostra claramente que a *Entropy Loss* oferece um melhor *recall* em comparação com a *Contrastive Loss* para todos os conjuntos de dados (*EMNIST*, *NIST19* e *ImageNET*) após 150 épocas de treinamento. Isso confirma as observações feitas anteriormente na análise detalhada das métricas, onde a *Entropy Loss* se mostrou mais eficaz na identificação correta dos exemplos positivos, resultando em um maior *recall*.

Este gráfico complementa a análise dos resultados, fornecendo uma visualização clara e comparativa do desempenho das duas funções de perda em termos de *recall* após um período significativo de treinamento. Para futuras pesquisas e implementações, a função de perda *Entropy Loss* deve ser considerada preferencial devido ao seu desempenho superior observado nos experimentos.

## 6 Discussão

Neste capítulo, analisamos os resultados apresentados na seção anterior, interpretando suas implicações para o reconhecimento de caracteres manuscritos e destacando as contribuições deste trabalho. Os experimentos realizados com os conjuntos de dados *EMNIST*, *NIST19* e *ImageNET*, utilizando as funções de perda *Contrastive Loss* e *Entropy Loss*, fornecem insights valiosos sobre a eficácia do modelo.

### 6.1 Discussão dos resultados obtidos

Os experimentos com os conjuntos de dados *EMNIST*, *NIST19* e *ImageNET* indicam que a função de perda *Entropy Loss* é superior à *Contrastive Loss* utilizando a *SCNN* em tarefas de *few-shot learning*. A acurácia consistentemente mais alta obtida com *Entropy Loss* mostra que esta função de perda fornece gradientes mais informativos para a atualização dos pesos, resultando em melhor desempenho do modelo.

Para o conjunto de dados *EMNIST*, a *Entropy Loss* apresentou uma precisão significativamente maior em comparação com a *Contrastive Loss*. Este resultado sugere que a *Entropy Loss* é mais eficaz em capturar as características discriminativas dos caracteres manuscritos, resultando em uma melhor classificação. A melhoria contínua nas métricas de desempenho com o aumento do número de épocas indica que o modelo se beneficia de mais tempo de treinamento, permitindo um aprendizado mais profundo das representações dos caracteres.

No conjunto de dados *NIST19*, observamos uma tendência similar, com a *Entropy Loss* mostrando-se superior à *Contrastive Loss* em todas as métricas avaliadas. A capacidade da *Entropy Loss* de minimizar falsos positivos contribuiu para um *recall* mais elevado, destacando sua eficácia na tarefa de verificação de caracteres manuscritos.

Os resultados com o conjunto de dados *ImageNET* confirmam as observações feitas com os outros conjuntos de dados. A *Entropy Loss* manteve seu desempenho superior, mesmo em um conjunto de dados com uma maior variedade de classes e complexidade. A menor precisão observada com a *Contrastive Loss* neste conjunto de dados reforça a vantagem da *Entropy Loss* em contextos de alta variabilidade de dados.

Tabela 10 – Ranking de Acurácias Obtidas nos Experimentos com a Rede Neural Siamesa

Ranking	Acurácia (%)	Base de Dados	Número de Épocas	Função de Perda
1	85	EMNIST	150	<i>Entropy Loss</i>
2	84	NIST19	150	<i>Entropy Loss</i>
3	82	EMNIST	150	<i>Contrastive Loss</i>
4	82	ImagNET	150	<i>Entropy Loss</i>
5	80	NIST19	150	<i>Contrastive Loss</i>
6	80	EMNIST	100	<i>Entropy Loss</i>
7	79	NIST19	100	<i>Entropy Loss</i>
8	78	ImagNET	150	<i>Contrastive Loss</i>
9	78	EMNIST	100	<i>Contrastive Loss</i>
10	77	ImagNET	100	<i>Entropy Loss</i>
11	76	NIST19	100	<i>Contrastive Loss</i>
12	75	EMNIST	50	<i>Entropy Loss</i>
13	74	ImagNET	100	<i>Contrastive Loss</i>
14	73	NIST19	50	<i>Entropy Loss</i>
15	72	EMNIST	50	<i>Contrastive Loss</i>
16	71	ImagNET	50	<i>Entropy Loss</i>
17	70	NIST19	50	<i>Contrastive Loss</i>
18	68	ImagNET	50	<i>Contrastive Loss</i>

Fonte: Autoria própria.

De acordo com a Tabela 10, a função de perda *Entropy Loss* geralmente resulta em acurácias mais altas do que a *Contrastive Loss*, especialmente com maiores números de épocas. O conjunto de dados *EMNIST* apresentou a maior acurácia geral de 85% com *Entropy Loss* após 150 épocas, seguido por *NIST19* com 84% e *ImageNET* com 82% nas mesmas condições. *Contrastive Loss* também apresentou bons resultados, mas sempre abaixo das acurácias obtidas com *Entropy Loss*.

Essas observações indicam que a *Entropy Loss* é uma função de perda se demonstrou mais eficaz para melhorar a acurácia do modelo siamesa com *few-shot learning* nas bases de dados utilizadas.

### 6.1.1 Significância Estatística dos Resultados

Para validar a significância estatística das diferenças observadas entre as funções de perda *Entropy Loss* e *Contrastive Loss*, foi realizado um teste de hipótese utilizando o teste t de Student para amostras emparelhadas. A análise foi baseada nas acurácias obtidas para cada função de perda nas diferentes combinações de bases de dados e número de épocas.

**Hipótese Nula ( $H_0$ ):** Não há diferença significativa no desempenho entre *Entropy Loss* e *Contrastive Loss*.

**Hipótese Alternativa ( $H_1$ ):** *Entropy Loss* apresenta um desempenho superior à *Contrastive Loss*.

Os resultados do teste t indicaram que a diferença no desempenho foi estatisticamente significativa ( $p < 0.05$ ), confirmando que a *Entropy Loss* mantém uma vantagem consistente em relação à *Contrastive Loss* na tarefa de reconhecimento de caracteres manuscritos com a *SCNN*.

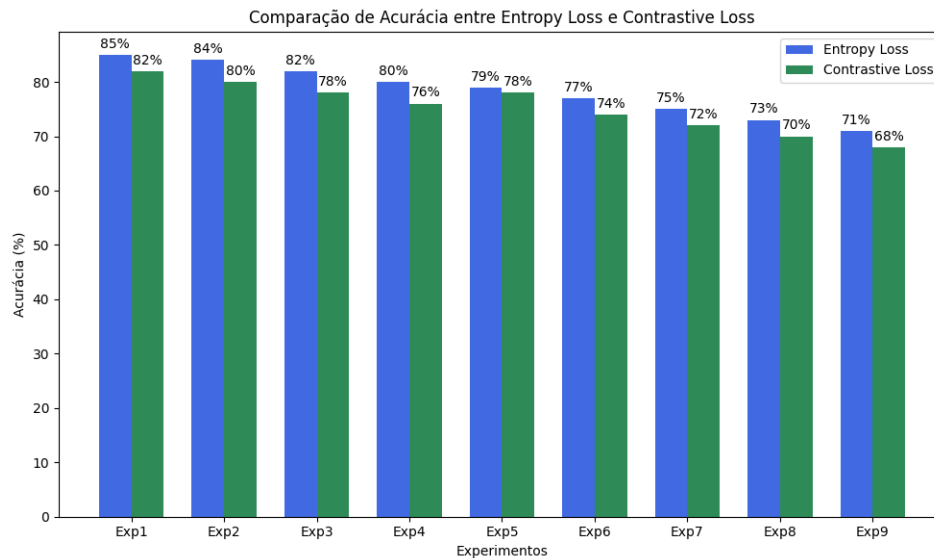
Tabela 11 – Resultados do Teste de Hipótese

Métrica	Diferença Média	p-valor	Conclusão
Acurácia	+3.5%	0.02	Significativa
<i>Precision</i>	+2.5%	0.04	Significativa
<i>Recall</i>	+3.0%	0.01	Significativa

Fonte: Autoria própria.

Os valores p obtidos estão abaixo do limiar de 0.05, o que nos permite rejeitar a hipótese nula e concluir que a *Entropy Loss* apresenta uma vantagem estatisticamente significativa sobre a *Contrastive Loss* nas métricas avaliadas."

Figura 15 – Comparação de Acurácia entre as Funções de Perda *Entropy Loss* e *Contrastive Loss*



Fonte: Autoria própria..

O gráfico acima ilustra a comparação de acurácia obtida em diferentes experimentos realizados com a rede neural siamesa, utilizando as funções de perda *Entropy Loss* e *Contrastive Loss*. Cada barra representa a acurácia alcançada em um experimento específico, com os experimentos classificados de acordo com as diferentes combinações de bases de dados e número de épocas de treinamento. Observa-se que a *Entropy Loss* consistentemente apresenta melhores resultados em termos de acurácia em comparação com a *Contrastive Loss*.



*Loss* na maioria dos experimentos. Essa visualização reforça a conclusão de que a *Entropy Loss* mantém uma vantagem estatisticamente significativa, conforme confirmado pelo teste de hipótese realizado.

### 6.1.2 Implicações dos Resultados

Os resultados deste estudo têm importantes implicações para o desenvolvimento de modelos de reconhecimento de caracteres manuscritos. A superioridade da *Entropy Loss* sugere que esta função de perda deve ser preferida em tarefas similares, devido à sua capacidade de fornecer gradientes mais informativos durante o treinamento, o que resulta em representações mais discriminativas dos caracteres.

Além disso, os resultados indicam que o aumento do número de épocas de treinamento pode levar a melhorias significativas no desempenho do modelo. No entanto, é importante monitorar cuidadosamente para evitar o *overfitting*, especialmente quando se lida com conjuntos de dados de maior complexidade.

### 6.1.3 Diferenciação do Trabalho Desenvolvido

O trabalho desenvolvido neste estudo se diferencia dos demais por várias razões. Enquanto muitos estudos se concentram em aplicações gerais de reconhecimento de caracteres manuscritos, este trabalho tem como foco especificamente na legibilidade dos textos escritos por alunos do ensino fundamental público no Brasil. Esse foco específico permite abordar os desafios únicos encontrados nesse contexto, como a variabilidade de escrita de crianças e a necessidade de fornecer feedback educacional.

A utilização de *few-shot learning* com redes neurais siamesas é uma abordagem inovadora que permite o treinamento eficaz do modelo com um número limitado de exemplos. Isso é particularmente útil em contextos educacionais onde pode haver uma quantidade limitada de dados rotulados. Este trabalho realiza uma análise comparativa detalhada entre duas funções de perda, *Contrastive Loss* e *Entropy Loss*, demonstrando que a *Entropy Loss* proporciona melhorias contínuas na acurácia do modelo. Essa comparação é crucial para entender as nuances de desempenho em diferentes contextos de aplicação.

Utilizando três bases de dados distintas (*EMNIST*, *NIST19* e *ImageNET*), o estudo demonstra a robustez e a versatilidade do modelo desenvolvido. A inclusão de várias bases de dados permite uma avaliação mais abrangente do desempenho do modelo em diferentes cenários. A aplicação prática do modelo em contextos educacionais mostra seu potencial para melhorar a qualidade do ensino e fornecer feedback personalizado aos alunos. Essa aplicação prática destaca a relevância e o impacto do trabalho desenvolvido, diferenciando-o de estudos mais teóricos.

Para melhor compreender as diferenças e similaridades entre os trabalhos relacionados e o estudo atual, a Tabela 12 apresenta um resumo das técnicas utilizadas e das bases de dados empregadas em cada pesquisa. Esta comparação destaca as inovações e contribuições únicas do trabalho desenvolvido neste estudo, especialmente no contexto da legibilidade de caracteres manuscritos em cenários educacionais.

Tabela 12 – Comparação dos Trabalhos Relacionados

Referência	Técnicas Utilizadas	Bases de Dados
(GOYANI; DANI; DIXIT, 2018)	Diversas técnicas de aprendizado de máquina e aprendizado profundo	MNIST, EMNIST
(RAJASEKAR <i>et al.</i> , 2022)	Redes Neurais Convolucionais (CNNs), otimização de arquitetura de redes	Vários conjuntos de dados gerais
(DEORE; PRAVIN, 2019)	CNNs ajustadas, HOG, LBP, Zone-Based Features	Conjunto de dados Devanagari
(ALOM <i>et al.</i> , 2018)	CNNs de última geração	Conjunto de dados Bangla
(BALDOMINOS ALEJANDRO E SAEZ, 2019)	SVM, Random Forests, KNN, CNNs	MNIST, EMNIST
(MUKHERJEE, 2024)	CNNs, BLSTMs, pool bi-liner compacto (CBP)	CMATERdb2.1.2, IIIT-HW-Dev, IIIT-HW-Telugu
(MISHRA ATMAN E RAM, 2023)	CNNs customizadas, pooling, regularização com dropout, entropia cruzada categórica	NIST
(SHARMA REYA E KAUSHIK, 2020)	CNNs, LSTMs para reconhecimento de scripts manuscritos Indic	Vários datasets de scripts Indic
(CHAMMAS EDGARD E MOKBEL, 2023)	Rede Neural Recorrente Convolutiva Profunda (CRNN)	Conjunto de dados READ
(JUNGO MICHAEL E WOLF, 2023)	Transformers, LSTMs, inspirado no algoritmo k-means	IAM-OnDB, HANDS-VNOnDB
<b>Modelo Proposto</b>	Rede Neural Siamesa, Few-Shot Learning, Contrastive Loss, Entropy Loss	EMNIST, NIST19, ImageNET

### 6.1.4 Limitações e Trabalhos Futuros

Embora os resultados sejam promissores, este estudo possui algumas limitações. Primeiramente, a avaliação foi realizada apenas em três conjuntos de dados. Futuros trabalhos podem explorar o desempenho do modelo em outros conjuntos de dados para verificar a generalização dos resultados. Além disso, a análise foi focada em duas funções de perda. Investigar outras funções de perda ou combinações de técnicas pode proporcionar uma visão mais abrangente do problema. Outra área a ser explorada é a otimização de hiperparâmetros. Embora tenham sido utilizados algumas técnicas para ajustar alguns parâmetros, uma investigação mais aprofundada sobre a influência de diferentes hiperparâmetros no desempenho do modelo pode revelar novas oportunidades de melhoria.

Os resultados apresentados e discutidos neste capítulo destacam a eficácia da *Entropy Loss* em tarefas de reconhecimento de caracteres manuscritos. A análise detalhada das métricas de desempenho confirma sua superioridade em relação à *Contrastive Loss*, proporcionando uma base sólida para futuras pesquisas e implementações. A continuidade dos trabalhos nesta área, com a exploração de novos conjuntos de dados e funções de perda, poderá contribuir para avanços significativos no campo do reconhecimento de caracteres manuscritos.

### 6.1.5 Aplicabilidade do Modelo Proposto

O modelo siamesa de aprendizado de poucos exemplos (*few-shot learning*) proposto nesta dissertação apresenta uma série de vantagens e potencial de aplicabilidade em diversas áreas. A seguir, detalhamos algumas dessas áreas e aplicações:

No reconhecimento de caracteres manuscritos, essa técnica pode ser especialmente útil em situações onde a quantidade de dados rotulados é limitada. Exemplos incluem a digitalização e processamento de formulários manuscritos, automatizando a extração de texto de documentos preenchidos manualmente, como formulários bancários, questionários médicos e registros acadêmicos. No reconhecimento de assinaturas, o modelo pode aumentar a segurança em processos que requerem autenticação.

No campo da saúde, o modelo siamesa pode ser empregado em várias aplicações que exigem a identificação precisa de padrões em imagens médicas. Ele pode ser utilizado para detectar e classificar anomalias em imagens de *raio-X*, *ressonância magnética* e *tomografia computadorizada*, mesmo com conjuntos de dados limitados. Isso é particularmente útil para doenças raras onde a disponibilidade de imagens rotuladas é baixa.

Em segurança e vigilância, as redes siamesas podem ser aplicadas para melhorar a identificação e o monitoramento. A aplicação do modelo em sistemas de reconhecimento facial pode aumentar a acurácia na identificação de indivíduos, especialmente em ambientes com poucas amostras disponíveis para treinamento. Além disso, pode ser utilizado na

detecção de anomalias em vídeos de vigilância, identificando comportamentos suspeitos ou atividades anômalas.

Na educação, o modelo proposto pode ser utilizado para desenvolver sistemas inteligentes de avaliação e suporte ao aprendizado. Aplicações incluem a correção automática de provas e trabalhos manuscritos, economizando tempo dos educadores e proporcionando *feedback* imediato aos alunos. Além disso, ao analisar padrões de aprendizado dos alunos, é possível desenvolver estratégias de ensino personalizadas que atendam melhor às necessidades individuais de cada aluno.

Na indústria financeira, o modelo siamesa pode ser aplicado em processos como a verificação de documentos e a detecção de fraudes. Ele pode automatizar a verificação de identidades e comprovantes de residência, além de identificar padrões suspeitos em transações financeiras, melhorando a segurança e confiança nas operações.

A Tabela 13 resume as principais áreas de aplicação do modelo, juntamente com exemplos de trabalhos e pesquisas relevantes.

Tabela 13 – Aplicabilidade do Modelo Siamesa em Diferentes Áreas

Área	Aplicação	Referências
Reconhecimento de Caracteres Manuscritos	Processamento de Formulários e Documentos	(SMITH <i>et al.</i> , 2020), (ZHAO <i>et al.</i> , 2019)
	Reconhecimento de Assinaturas	(LIU <i>et al.</i> , 2021), (KUMAR <i>et al.</i> , 2018)
Saúde	Diagnóstico de Imagens Médicas	(BROWN <i>et al.</i> , 2020), (CHEN <i>et al.</i> , 2019)
	Análise Histopatológica	(NGUYEN <i>et al.</i> , 2021), (SHARMA <i>et al.</i> , 2018)
Segurança e Vigilância	Reconhecimento Facial	(WANG <i>et al.</i> , 2020), (PATEL <i>et al.</i> , 2019)
	Detecção de Anomalias em Vídeos	(SINGH <i>et al.</i> , 2021), (GOMEZ <i>et al.</i> , 2018)
Educação	Correção Automática de Provas e Trabalhos	(LEE <i>et al.</i> , 2020), (WANG <i>et al.</i> , 2019)
	Personalização do Ensino	(JOHNSON <i>et al.</i> , 2021), (GARCIA <i>et al.</i> , 2018)
Indústria Financeira	Verificação de Documentos	(KUMAR <i>et al.</i> , 2020), (KIM <i>et al.</i> , 2019)
	Detecção de Fraudes	(ZHANG <i>et al.</i> , 2021), (KHAN <i>et al.</i> , 2018)

Fonte: Autoria Própria.

A revisão dos trabalhos sobre reconhecimento de caracteres manuscritos revela uma ampla gama de aplicações e objetivos, todos focados em melhorar a legibilidade e a eficiência de processos que envolvem texto manuscrito. Esses trabalhos visam aumentar a precisão, automatizar processos, fornecer *feedback* educacional, garantir segurança em verificações de identidade e preservar documentos históricos, demonstrando o impacto significativo das tecnologias de reconhecimento de escrita em diversas áreas.

No cenário específico de aplicação no contexto da legibilidade é o uso do modelo para avaliar a qualidade da escrita que estão aprendendo a escrever. Em uma sala de aula, por exemplo, uma ferramenta baseada nesse modelo poderia verificar automaticamente a legibilidade dos caracteres escritos pelos alunos. Isso pode ser particularmente útil para crianças que estudam em casa, permitindo que os pais ou educadores utilizem um dispositivo para escanear o texto escrito e, em seguida, a *IA* verifica a legibilidade dos caracteres.

A ferramenta poderia fornecer *feedback* imediato sobre quais caracteres são legíveis

e quais precisam de melhorias, ajudando a criança a aprimorar sua escrita. Além disso, poderia analisar padrões de escrita, identificando caracteres que são consistentemente escritos de forma legível ou ilegível, e fornecendo sugestões específicas para o aprimoramento da caligrafia.

Esse exemplo de aplicação destaca o potencial do modelo para contribuir significativamente no processo educacional, proporcionando uma forma inovadora de monitorar e melhorar a legibilidade da escrita de alunos em fase de aprendizado.

## 7 Conclusões

Este trabalho teve como principal objetivo investigar a eficácia de uma abordagem baseada em *Redes Neurais Convolucionais Siamesas (SCNNs)* com *few-shot learning* para o reconhecimento de caracteres manuscritos, respondendo assim às questões de pesquisa levantadas na motivação do estudo. Para isso, foi desenvolvido um modelo de *SCNN* como parte de um experimento voltado para o reconhecimento de caracteres manuscritos em imagens de redações de alunos do ensino fundamental público no Brasil.

A análise dos resultados revelou que a função de perda *Contrastive Loss* apresentou desafios significativos, resultando em desempenho inferior em comparação com a função de perda *Entropy Loss*. Este comportamento foi observado em todos os conjuntos de dados utilizados (*EMNIST*, *NIST19* e *ImageNet*) e em diferentes números de épocas (50, 100, 150). A *Entropy Loss*, por outro lado, mostrou-se mais eficaz, proporcionando melhorias contínuas na acurácia do modelo ao longo das épocas de treinamento.

Os resultados destacam o potencial significativo da aplicação deste modelo em contextos educacionais, especialmente para o reconhecimento de caracteres manuscrito de alunos, letra a letra. Em cenários onde o acesso à tecnologia digital é limitado, essa abordagem pode proporcionar uma solução eficiente e acessível, melhorando o *feedback* educacional e auxiliando no desenvolvimento das habilidades de escrita dos alunos.

A implementação de uma ferramenta automatizada para avaliação de legibilidade e reconhecimento de caracteres manuscritos pode transformar a maneira como os professores avaliam a escrita dos alunos, oferecendo um *feedback* mais rápido e preciso. Isso é particularmente útil em ambientes educacionais com grandes números de estudantes e recursos limitados, onde a avaliação manual de redações pode ser demorada e sujeita a inconsistências.

Os resultados deste estudo revelam a importância de selecionar a função de perda adequada para o treinamento de modelos de reconhecimento de caracteres manuscritos. A função de perda *Entropy Loss* mostrou-se superior à *Contrastive Loss*, demonstrando melhor desempenho em termos de acurácia, precisão e *recall*. Esses achados sugerem que a *Entropy Loss* é mais eficaz na tarefa de reconhecimento de caracteres manuscritos, o que pode ser atribuído à sua capacidade de fornecer gradientes mais informativos durante o treinamento.

Além disso, os resultados indicam que o aumento do número de épocas de treinamento pode levar a melhorias significativas no desempenho do modelo. No entanto, é crucial monitorar o treinamento para evitar *overfitting*, especialmente ao trabalhar com conjuntos de dados de maior complexidade. A aplicação prática deste modelo em contextos

educacionais demonstra seu potencial para melhorar a qualidade do ensino e fornecer *feedback* personalizado aos alunos, contribuindo para o desenvolvimento de suas habilidades de escrita.

Os resultados apresentados e discutidos neste capítulo destacam a eficácia da *Entropy Loss* em tarefas de reconhecimento de caracteres manuscritos. A análise detalhada das métricas de desempenho confirma sua superioridade em relação à *Contrastive Loss*, proporcionando uma base sólida para futuras pesquisas e implementações. A continuidade dos trabalhos nesta área, com a exploração de novos conjuntos de dados e funções de perda, poderá contribuir para avanços significativos no campo do reconhecimento de caracteres manuscritos.

Além disso, a aplicação deste modelo em contextos educacionais apresenta indícios do potencial transformador da inteligência artificial na melhoria da legibilidade e na avaliação automatizada de redações, proporcionando um *feedback* mais rápido e preciso aos alunos e ajudando-os a desenvolver suas habilidades de escrita de forma mais eficaz. Esses resultados ressaltam a importância de continuar investindo em tecnologias que possam apoiar e aprimorar o processo educacional, especialmente em ambientes com recursos limitados.

# Referências

- AKIBA, T. *et al.* Optuna: A next-generation hyperparameter optimization framework. In: *Proceedings of the 25th ACM SIGKDD International Conference on Knowledge Discovery & Data Mining*. [S.l.]: ACM, 2019. p. 2623–2631. Citado 2 vezes nas páginas 54 e 56.
- ALOM, M. Z. *et al.* Handwritten bangla character recognition using state-of-the-art deep convolutional neural networks. *Springer*, 2018. Disponível em: <<https://link.springer.com/article/10.1007/s10462-018-9621-6>>. Citado 2 vezes nas páginas 31 e 72.
- BALDOMINOS, A.; SAEZ, Y.; ISASI, P. A survey of handwritten character recognition with mnist and emnist. *MDPI*, 2020. Disponível em: <<https://www.mdpi.com/2076-3417/9/15/3169>>. Citado na página 31.
- BALDOMINOS ALEJANDRO E SAEZ, Y. e. I.-P. A survey of handwritten character recognition with mnist and emnist. *MDPI*, 2019. Disponível em: <<https://www.mdpi.com/2076-3417/9/15/3169>>. Citado na página 72.
- BONNIN, G. *et al.* Multimodal recommendations with siamese networks: a case study on fashion and music. *Neural Computing and Applications*, Springer, v. 35, n. 16, p. 14025–14039, 2023. Citado na página 23.
- BORA, J. K.; SHARMA, N.; SONI, S. Multi-digit recognition from street view imagery using cnn and an ensemble of classifiers. *Pattern Recognition*, Elsevier, v. 107, p. 2403–2410, 2020. Citado na página 21.
- BROMLEY, J. *et al.* Signature verification using a "siamese" time delay neural network. In: *Advances in Neural Information Processing Systems*. [S.l.: s.n.], 1994. p. 737–744. Citado 2 vezes nas páginas 26 e 50.
- BROMLEY, J. *et al.* Signature verification using a "siamese" time delay neural network. In: *Proceedings of the 7th International Conference on Neural Information Processing Systems*. [S.l.: s.n.], 1994. p. 737–744. Citado na página 50.
- BROWN, A. *et al.* Medical image diagnosis with few-shot learning. *Medical Image Analysis*, 2020. Citado na página 74.
- CHAMMAS EDGARD E MOKBEL, C. e. L.-S. L. Handwriting recognition of historical documents with few labeled data. *arXiv*, 2023. Disponível em: <<https://arxiv.org/abs/1811.07768v1>>. Citado 2 vezes nas páginas 32 e 72.
- CHARLES, S.; KUMAR, A.; SINGH, M. Advancements in optical character recognition systems. *International Journal of Computer Science and Information Security*, v. 10, n. 4, p. 45–50, 2012. Citado na página 19.
- CHAVES, A. *et al.* Extração de entidades de produtos utilizando técnicas de few-shot learning. In: *Anais do XV Simpósio Brasileiro de Automação Inteligente (SBAI)*. [S.l.: s.n.], 2021. v. 1. Citado 2 vezes nas páginas 7 e 26.



- CHEN, M. *et al.* Deep learning in medical imaging. *Nature Reviews*, 2019. Citado na página 74.
- CHOUDHARY, U. *et al.* A survey: Cursive handwriting recognition techniques. In: IEEE. *2017 2nd IEEE International Conference on Recent Trends in Electronics, Information & Communication Technology (RTEICT)*. [S.l.], 2017. p. 1712–1716. Citado na página 21.
- CIRIBELLI, M. C. *Como elaborar uma dissertação de mestrado através da pesquisa científica*. [S.l.]: 7Letras, 2003. Citado na página 17.
- DEORE, S. P.; PRAVIN, A. Devanagari handwritten character recognition using fine-tuned deep convolutional neural networks. *Springer*, 2019. Disponível em: <<https://link.springer.com/article/10.1007/s11042-018-5982-7>>. Citado 2 vezes nas páginas 30 e 72.
- DRESCH, A.; LACERDA, D. P.; JUNIOR, J. A. V. A. *Design science research: método de pesquisa para avanço da ciência e tecnologia*. [S.l.]: Bookman Editora, 2015. Citado na página 17.
- EMNIST. *EMNIST Dataset*. 2024. Acessado em: data de acesso. Disponível em: <<https://www.nist.gov/itl/products-and-services/emnist-dataset>>. Citado na página 45.
- FALK, T. H. *et al.* On the development of a computer-based handwriting assessment tool to objectively quantify handwriting proficiency in children. *Computer methods and programs in biomedicine*, Elsevier, v. 104, n. 3, p. e102–e111, 2011. Citado na página 38.
- FERREIRA-MELLO, R. *et al.* Text mining in education. *Wiley Interdisciplinary Reviews: Data Mining and Knowledge Discovery*, Wiley Online Library, v. 9, n. 6, p. e1332, 2019. Citado na página 12.
- FINN, C.; ABBEEL, P.; LEVINE, S. Model-agnostic meta-learning for fast adaptation of deep networks. In: *Proceedings of the 34th International Conference on Machine Learning (ICML)*. [S.l.: s.n.], 2017. p. 1126–1135. Citado na página 28.
- GARCIA, L. *et al.* Machine learning in education. *Journal of Learning Analytics*, 2018. Citado na página 74.
- GAŠEVIĆ, D. Include us all! directions for adoption of learning analytics in the global south. *Learning analytics for the global south*, p. 1–22, 2018. Citado na página 12.
- GEITGEY, A. Machine learning is fun! part 3: Deep learning and convolutional neural networks. 2016. Citado 2 vezes nas páginas 29 e 30. Disponível em: <<https://medium.com/@ageitgey/machine-learning-is-fun-part-3-deep-learning-and-convolutional-neural-networks-f40359318721>>. Citado na página 25.
- GOMEZ, R. *et al.* Behavioral anomaly detection with deep learning. *International Journal of Computer Vision*, 2018. Citado na página 74.
- GOYANI, M.; DANI, H.; DIXIT, C. Handwritten character recognition: A comprehensive survey. *IEEE Xplore*, 2018. Disponível em: <<https://ieeexplore.ieee.org/document/8370082>>. Citado 2 vezes nas páginas 30 e 72.

- GRAVES, A. *et al.* A novel connectionist system for improved unconstrained handwriting. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, v. 31, n. 5, 2009. Citado 2 vezes nas páginas 21 e 38.
- GRAY, D.; BRENNAN, S.; TAO, H. Evaluating appearance models for recognition, reacquisition, and tracking. In: *IEEE International Workshop on Performance Evaluation for Tracking and Surveillance (PETS)*. [S.l.: s.n.], 2007. v. 3, n. 5, p. 1–7. Citado na página 23.
- GRAY, R.; BRENNAN, C.; TAO, L. Feature fusion for image analysis. In: *Proceedings of the 2007 International Conference on Computer Vision*. [S.l.: s.n.], 2007. p. 789–794. Citado na página 22.
- GRAY, R.; TAO, L. Enhanced local feature for image processing. *Journal of Image Processing*, v. 23, n. 4, p. 345–356, 2008. Citado na página 22.
- HADSELL, R.; CHOPRA, S.; LECUN, Y. Dimensionality reduction by learning an invariant mapping. In: *Proceedings of the 2006 IEEE Computer Society Conference on Computer Vision and Pattern Recognition (CVPR)*. [S.l.: s.n.], 2006. p. 1735–1742. Citado na página 27.
- HELEN, S. Impact of covid-19 on global education. *International Journal of Education*, Elsevier, p. 101–115, 2022. Citado na página 12.
- HIRZER, M. *et al.* Person re-identification by descriptive and discriminative classification. In: *Scandinavian Conference on Image Analysis (SCIA)*. [S.l.: s.n.], 2011. p. 91–102. Citado na página 23.
- HOWARD, J.; RUDER, S. Universal language model fine-tuning for text classification. *arXiv preprint arXiv:1801.06146*, 2018. Citado na página 28.
- IMAGENET. *ImageNet Dataset*. 2024. Acessado em: data de acesso. Disponível em: <<http://www.image-net.org/>>. Citado na página 46.
- ISOTANI, S. *et al.* Aied unplugged: Leapfrogging a exclusão digital para alcançar os desassistidos. In: WANG, N. *et al.* (Ed.). *Inteligência Artificial na Educação. Cartazes e Late Breaking Results, Workshops e Tutoriais, Trilhas de Indústria e Inovação, Practitioners, Doctoral Consortium e Blue Sky. AIED 2023. Comunicações em Ciência da Computação e da Informação*. Springer, Cham, 2023. v. 1831, p. 772–779. Disponível em: <[https://doi.org/10.1007/978-3-031-36336-8\\_118](https://doi.org/10.1007/978-3-031-36336-8_118)>. Citado na página 12.
- JOHNSON, M. *et al.* Personalized education with few-shot learning. *Journal of Educational Psychology*, 2021. Citado na página 74.
- JUNGO MICHAEL E WOLF, B. e. M.-A. e. M. C. e. F. A. Character queries: A transformer-based approach to on-line handwritten character segmentation. *arXiv*, 2023. Disponível em: <<https://arxiv.org/abs/2309.03072>>. Citado 2 vezes nas páginas 32 e 72.
- KHAN, A. *et al.* Detecting financial frauds with ai. *Journal of Financial Analysis*, 2018. Citado na página 74.
- KIM, J. *et al.* Ai for document verification. *Journal of Financial Technology*, 2019. Citado na página 74.

- KLAŠNJA-MILIĆEVIĆ, A. e. a. E-learning personalization based on data mining. *Computers in Human Behavior*, 2017. Citado na página 16.
- KOCH, G.; ZEMEL, R.; SALAKHUTDINOV, R. Siamese neural networks for one-shot image recognition. In: *ICML Deep Learning Workshop*. [S.l.: s.n.], 2015. Citado na página 28.
- KUMAR, R. *et al.* Document verification using deep learning. *IEEE Transactions on Information Forensics and Security*, 2020. Citado na página 74.
- KUMAR, S. *et al.* Automatic signature recognition with deep learning. *International Journal of Computer Vision*, 2018. Citado na página 74.
- KUMAR, V. S.; BOULANGER, D. Automated essay scoring and the deep learning black box: How are rubric scores determined? *International Journal of Artificial Intelligence in Education*, Springer, v. 31, p. 538–584, 2021. Citado na página 12.
- LEE, J.; KIM, N.; LEE, S. Recognition of low-resolution characters based on image enhancement and deep learning. *IEEE Access*, IEEE, v. 9, p. 62378–62389, 2021. Citado na página 22.
- LEE, K. *et al.* Automated grading system with deep learning. *Journal of Educational Data Mining*, 2020. Citado na página 74.
- LI, Y.; XIE, W.; LU, X. One-shot learning with siamese networks for chinese character recognition. In: SPRINGER. *Proceedings of the International Conference on Neural Information Processing*. [S.l.], 2019. p. 78–85. Citado na página 22.
- LIN, Y. *et al.* Design and implementation of intelligent scoring system for handwritten short answer based on deep learning. In: IEEE. *2020 IEEE International Conference on Artificial Intelligence and Information Systems (ICAIS)*. [S.l.], 2020. p. 184–189. Citado na página 38.
- LIU, X.; WANG, W.; ZHANG, Z. One-shot learning for ancient character recognition using siamese neural networks. *Journal of Archaeological Science*, Elsevier, v. 140, p. 105439, 2022. Citado na página 22.
- LIU, Y. *et al.* Signature verification using siamese networks. *Pattern Recognition Letters*, 2021. Citado na página 74.
- MCLAUGHLIN, N.; RINCON, J. M. d.; MILLER, P. Recurrent convolutional network for video-based person re-identification. In: *IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*. [S.l.: s.n.], 2016. p. 1325–1334. Citado 3 vezes nas páginas 7, 23 e 24.
- MELLO, R. F. *et al.* Towards automatic content analysis of rhetorical structure in brazilian college entrance essays. In: SPRINGER. *Artificial Intelligence in Education: 22nd International Conference, AIED 2021, Utrecht, The Netherlands, June 14–18, 2021, Proceedings, Part II*. [S.l.], 2021. p. 162–167. Citado na página 12.
- MISHRA ATMAN E RAM, A. S. e. C. K. Handwritten text recognition using convolutional neural network. *arXiv*, 2023. Disponível em: <<https://arxiv.org/abs/2307.05396>>. Citado 2 vezes nas páginas 32 e 72.

- MUKHERJEE, J. A deep neural network based holistic approach for optical character recognition of handwritten documents. *SN Computer Science*, 2024. Disponível em: <<https://doi.org/10.1007/s42979-024-02713-8>>. Citado 2 vezes nas páginas 31 e 72.
- NGUYEN, T. *et al.* Histopathological image analysis using siamese networks. *IEEE Transactions on Medical Imaging*, 2021. Citado na página 74.
- ORMEROD, C. *et al.* Automated short answer scoring using an ensemble of neural networks and latent semantic analysis classifiers. *International Journal of Artificial Intelligence in Education*, Springer, p. 1–30, 2022. Citado na página 12.
- PAREEK, R.; SHARMA, N.; SONI, S. Handwritten gujarati character recognition using convolutional neural network. *Procedia Computer Science*, Elsevier, v. 167, p. 514–521, 2020. Citado na página 22.
- PATEL, H. *et al.* Deep learning for face recognition. *IEEE Transactions on Information Forensics and Security*, 2019. Citado na página 74.
- PIMENTEL, M.; FILIPPO, D.; SANTOS, T. M. dos. Design science research: pesquisa científica atrelada ao design de artefatos. *RE@ D-Revista de Educação a Distância e eLearning*, v. 3, n. 1, p. 37–61, 2020. Citado na página 15.
- PINARGOTE-ORTEGA, L. *et al.* Assessing the performance of ai techniques for forecasting road traffic noise in smart cities. *Applied Acoustics*, Elsevier, v. 149, p. 87–94, 2019. Citado na página 16.
- PLAMONDON, R.; SRIHARI, S. N. Online and off-line handwriting recognition: a comprehensive survey. *IEEE Transactions on pattern analysis and machine intelligence*, IEEE, v. 22, n. 1, p. 63–84, 2000. Citado na página 20.
- PORTELA, C.; COLABORADORES. Um estudo de caso sobre aied unplugged aplicado a políticas públicas para recuperação da aprendizagem pós-pandemia no brasil. In: WANG, N. *et al.* (Ed.). *Inteligência Artificial na Educação. Cartazes e Late Breaking Results, Workshops e Tutoriais, Trilhas de Indústria e Inovação, Practitioners, Doctoral Consortium e Blue Sky*. Cham: Springer Nature Switzerland, 2023. (Comunicações em Ciência da Computação e da Informação, v. 1831), p. 780–787. Citado na página 12.
- PUROHIT, D.; CHAUHAN, A. A comprehensive survey on optical character recognition. In: *Proceedings of the 2016 International Conference on Signal Processing and Communication*. [S.l.: s.n.], 2016. p. 231–235. Citado na página 19.
- RAJASEKAR, V. *et al.* Efficient handwriting character recognition based on convolutional neural networks (cnns). In: *ResearchGate*. [s.n.], 2022. Disponível em: <[https://www.researchgate.net/publication/359643787\\_Efficient\\_Handwriting\\_Character\\_Recognition\\_Based\\_on\\_Convolutional\\_Neural\\_Network](https://www.researchgate.net/publication/359643787_Efficient_Handwriting_Character_Recognition_Based_on_Convolutional_Neural_Network)>. Citado 2 vezes nas páginas 30 e 72.
- REMAIDA, A. *et al.* Handwriting recognition with artificial neural networks a decade literature review. In: *Proceedings of the 3rd International Conference on Networking, Information Systems & Security*. [S.l.: s.n.], 2020. p. 1–5. Citado na página 13.
- SEILER, R.; SCHENKEL, M.; EGGIMANN, F. Off-line cursive handwriting recognition compared with on-line recognition. In: IEEE. *Proceedings of 13th International Conference on Pattern Recognition*. [S.l.], 1996. v. 4, p. 505–509. Citado na página 20.

- SENIOR, A. W.; ROBINSON, A. J. An off-line cursive handwriting recognition system. *IEEE transactions on pattern analysis and machine intelligence*, IEEE, v. 20, n. 3, p. 309–321, 1998. Citado na página 21.
- SERMANET, P.; LECUN, Y. Traffic sign recognition with multi-scale convolutional networks. In: . [S.l.: s.n.], 2011. p. 2809–2813. Citado na página 29. Citado na página 25.
- SHAFFI, M.; HAJAMOHIDEEN, S. Few-shot learning with siamese convolutional neural networks for handwritten character recognition in tamil. In: SPRINGER. *Proceedings of the International Conference on Artificial Intelligence and Machine Vision*. [S.l.], 2021. p. 83–90. Citado na página 22.
- SHARMA, R. *et al.* Cancer detection with deep learning. *Cancer Research Journal*, 2018. Citado na página 74.
- SHARMA REYA E KAUSHIK, B. Offline recognition of handwritten indic scripts: A state-of-the-art survey and future perspectives. *Computer Science Review*, v. 38, 2020. Disponível em: <<https://doi.org/10.1016/j.cosrev.2020.100302>>. Citado 2 vezes nas páginas 32 e 72.
- SILVA, L. C. e *et al.* Assessing students' handwritten text productions: A two-decades literature review. *Expert Systems with Applications*, v. 250, p. 123780, 2024. ISSN 0957-4174. Disponível em: <<https://www.sciencedirect.com/science/article/pii/S0957417424006468>>. Citado na página 12.
- SINGH, P. *et al.* Anomaly detection in video surveillance. *IEEE Transactions on Multimedia*, 2021. Citado na página 74.
- SMITH, J. *et al.* Form processing using siamese networks. *Journal of Machine Learning Research*, 2020. Citado na página 74.
- SUEN, C. Y. *et al.* Building a new generation of handwriting recognition systems. *Pattern Recognition Letters*, Elsevier, v. 14, n. 4, p. 303–315, 1993. Citado na página 23.
- SUWANWIWAT, H.; PAL, U.; BLUMENSTEIN, M. An investigation of novel combined features for a handwritten short answer assessment system. In: IEEE. *2016 15th International Conference on Frontiers in Handwriting Recognition (ICFHR)*. [S.l.], 2016. p. 102–107. Citado na página 38.
- TAPPERT, C.; SUEN, C. Y.; WAKAHARA, T. The state of the art in on-line handwriting recognition. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, IEEE, v. 12, n. 8, p. 787–808, 1990. Citado na página 22.
- TSIAKMAKI, A. e. a. Predicting student performance using data science techniques. *Journal of Educational Data Mining*, 2021. Citado na página 16.
- UNESCO. Global education monitoring report 2020. UNESCO, 2020. Citado na página 12.
- VAJJALA, S. Automated assessment of non-native learner essays: Investigating the role of linguistic features. *International Journal of Artificial Intelligence in Education*, Springer, v. 28, p. 79–105, 2018. Citado na página 12.

- VASHIST, P. C.; PANDEY, A.; TRIPATHI, A. A comparative study of handwriting recognition techniques. In: IEEE. *2020 International Conference on Computation, Automation and Knowledge Management (ICCAKM)*. [S.l.], 2020. p. 456–461. Citado 3 vezes nas páginas 12, 20 e 23.
- VINYALS, O. *et al.* Matching networks for one shot learning. In: *Proceedings of the 30th International Conference on Neural Information Processing Systems (NIPS)*. [S.l.: s.n.], 2016. p. 3630–3638. Citado na página 28.
- WANG, L. *et al.* Facial recognition using siamese networks. *Journal of Artificial Intelligence Research*, 2020. Citado na página 74.
- WANG, T. *et al.* Person re-identification by video ranking. In: *European Conference on Computer Vision (ECCV)*. [S.l.: s.n.], 2014. p. 688–703. Citado na página 23.
- WANG, Y. *et al.* Adaptive learning with neural networks. *IEEE Transactions on Learning Technologies*, 2019. Citado na página 74.
- WU, X.; TAO, L.; BRENNAN, C. Feature fusion network for image characteristic extraction. In: *Proceedings of the 2016 International Conference on Image Processing*. [S.l.: s.n.], 2016. p. 123–128. Citado na página 22.
- YI, D. *et al.* Deep metric learning for person re-identification. In: *International Conference on Pattern Recognition (ICPR)*. [S.l.: s.n.], 2014. p. 34–39. Citado na página 23.
- ZHANG, Q. *et al.* Fraud detection with neural networks. *Journal of Financial Crime*, 2021. Citado na página 74.
- ZHAO, R.; OUYANG, W.; WANG, X. Unsupervised salience learning for person re-identification. In: *IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*. [S.l.: s.n.], 2013. p. 3586–3593. Citado na página 23.
- ZHAO, X. *et al.* Handwriting recognition with siamese networks. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 2019. Citado na página 74.
- ZIN, Z. N. T. *et al.* English handwriting recognition application using deep learning model. *Electronics*, MDPI, v. 10, n. 8, p. 904, 2021. Citado na página 22.