INSTITUTO
UNIVERSITÁRIO
DE LISBOA

MANCAPT - Identificação de Autores e Transcrição de Manuscritos

Ednilson Fonseca Queni

Mestrado em Tecnologias Digitais para o Negócio

Orientador:
Professor Doutor João Carlos Amaro Ferreira,
Professor Auxiliar com Agregação,
ISCTE-IUL - Universidade de Lisboa

Coorientador:
Professor Doutor Luís Manuel Nobre de Brito Elvas,
Professor Auxiliar,
ISCTE-IUL - Universidade de Lisboa

Janeiro, 2025

TECNOLOGIAS
E ARQUITETURA

MANCAPT - Identificação de Autores e Transcrição de Manuscritos

Ednilson Fonseca Queni

Mestrado em Tecnologias Digitais para o Negócio

Orientador:
Professor Doutor João Carlos Amaro Ferreira,
Professor Auxiliar com Agregação,
ISCTE-IUL - Universidade de Lisboa

Coorientador:
Professor Doutor Luís Manuel Nobre de Brito Elvas,
Professor Auxiliar,
ISCTE-IUL - Universidade de Lisboa

Janeiro, 2025

Agradecimento

A concretização deste trabalho representa não só um marco académico, mas também uma jornada de crescimento pessoal e profissional, que só foi possível graças ao apoio e incentivo de diversas pessoas.

Em primeiro lugar, expresso a minha mais profunda gratidão à minha família, pelo suporte incondicional, pelo carinho e pela paciência demonstrada ao longo deste percurso. Aos meus pais, que sempre me incentivaram a perseguir os meus objetivos com determinação e resiliência. Um agradecimento especial à minha mãe, Deoclesiana Augusta Fonseca, a principal responsável pelo meu crescimento e sucesso, cujo amor inabalável, dedicação incansável e ensinamentos foram elementares para cada conquista da minha vida. A sua força e exemplo moldaram-me como pessoa e profissional. À minha mãe de coração, Ana Lopes, cujo apoio e carinho incondicionais foram fundamentais nesta jornada, o meu sincero reconhecimento.

Um agradecimento especial aos meus orientadores, o Professor Doutor João Carlos Amaro Ferreira e o Professor Doutor Luís Manuel Nobre de Brito Elvas, por toda a dedicação, pela partilha de conhecimento e pela orientação valiosa que contribuíram significativamente para a realização deste trabalho. O vosso apoio académico e as vossas críticas construtivas foram a chave para o desenvolvimento desta investigação.

A todos os colegas e amigos que de alguma forma me acompanharam e contribuíram para este processo, seja através de debates, troca de ideias ou simplesmente pelo apoio emocional, deixo o meu sincero agradecimento.

Por fim, às instituições e entidades que forneceram suporte académico e técnico, permitindo o desenvolvimento deste estudo, a minha gratidão. Sem o contributo de cada um, este trabalho não teria sido possível.

Resumo

Este trabalho explora os desafios da digitalização de documentos manuscritos, com foco na variabilidade da caligrafia e nas limitações dos sistemas de Reconhecimento Ótico de Caracteres (OCR) genéricos. Este trabalho consiste na criação de um protótipo designado por MANCAPT, que apresenta uma solução baseada em tecnologias avançadas, incluindo o *Microsoft AI Document Intelligence*, utilizado para a criação de modelos personalizados de diferentes autorias e transcrição dos manuscritos, e a arquitetura neural *InceptionV3*, utilizada para a identificação dos autores dos manuscritos com elevada precisão.

O MANCAPT integra técnicas de pré-processamento como a binarização, remoção de ruído e normalização da escala, complementadas por *data augmentation* para criar um conjunto de treino consistente. Foram utilizadas imagens manuscritas e impressas numa proporção equilibrada (70% impressas 30% manuscritas), permitindo ao modelo capturar tanto padrões claros como nuances estilísticas da escrita manual.

Os resultados experimentais demonstraram que a integração do *InceptionV3* e do *Azure AI Document Intelligence* no sistema aumentou a precisão da identificação de autores para 98% e reduziu a Taxa de Erro de Caracteres (CER) para 6%, em comparação com o modelo genérico do *Azure AI Document Intelligence*, que obteve um CER de 11,3% na transcrição de manuscritos. A abordagem personalizada do MANCAPT revelou-se crucial para lidar com as idiossincrasias da escrita, promovendo a eficiência em soluções de OCR.

Palavras-chave: Inteligência artificial, aprendizagem máquina, reconhecimento ótico de caracteres, identificação de autor.

Abstract

This work explores the challenges of digitizing handwritten documents, focusing on handwriting variability and the limitations of generic Optical Character Recognition (OCR) systems. This study involves the creation of a prototype named MANCAPT, which provides a solution based on advanced technologies, including Microsoft AI Document Intelligence for the development of customized models for different authors and the transcription of manuscripts, as well as the neural architecture InceptionV3 for high-precision author identification.

MANCAPT integrates preprocessing techniques such as binarization, noise removal, and scale normalization, complemented by data augmentation to create a robust and resilient training dataset. Handwritten and printed images were used in a balanced proportion (70% printed, 30% handwritten), allowing the model to capture both clear patterns and stylistic nuances of manual writing.

Experimental results demonstrated that integrating InceptionV3 with Azure AI Document Intelligence improved author identification accuracy to 98% and reduced the Character Error Rate (CER) to 6%, compared to the generic Azure AI Document Intelligence model, which achieved a CER of 11.3% in manuscript transcription. The personalized approach of MANCAPT proved crucial in addressing the idiosyncrasies of handwriting, enhancing the efficiency of OCR solutions.

Keywords: Artificial intelligence, machine learning, optical character recognition, author identification.

Índice

Agradecimento iii
Resumo v
Abstract vii
Índice ix
Índice de Figuras xiii
Índice de Tabelas xv
Lista Abreviaturas xvii
CAPÍTULO 1 1
Introdução 1
1.1. Motivação 1
1.2. Contexto 1
1.2.1 Transição do Analógico para o Digital 1
1.2.2 Importância da Escrita Manual na Era Digital 2
1.2.3 Importância da Identificação do Autor 3
1.3. Objetivos 3
1.4. Metodologia 4
1.5. Estrutura do Documento 5
CAPíTULO 2 7
Estado da Arte 7
2.1. Introdução ao Reconhecimento Ótico de Caracteres (OCR) 7
2.2. Tecnologias Utilizadas com OCR 8
2.2.1 Redes Neurais Convolucionais (CNN) 8
2.2.2 Redes Neurais Recorrentes (RNN) e Long Short-Term Memory (LSTM) 8
2.2.3 Modelos Ocultos de Markov (HMM) e Máquinas de Vetores de Suporte (SVM) 9
2.2.4 Técnicas de Pré-processamento 10
2.3. Identificação do Autor 11
2.3.1 Métodos Baseados em CNN e RNN 11
2.3.2 Análise Idiossincrática 11
2.3.3 Métodos de Identificação de Autores 12
2.4. Síntese Crítica da Literatura 13
2.5. Discussão de Resultados 15
2.6. Conclusão 17
CAPÍTULO 3 19

MANCAPT 19
3.1. Requisitos 19
3.2. Desenho da Arquitetura 21
3.2.1 Arquitetura Geral do MANCAPT 21
3.2.2 Identificação do Autor com *InceptionV3* 21
3.2.3 Transcrição de Manuscritos com *Azure AI Document Intelligence* 23
CAPÍTULO 4 25
Implementação 25
4.1. Recolha de Amostras 26
4.2. Tratamento de Dados 26
4.3. Processo de Identificação do Autor 27
4.3.1 Pré-processamento e *Data Augmentation* 28
4.3.2 Treino do Modelo 28
4.3.2 Ajuste de Hiperparâmetros 29
4.3.4 Resultados do Treino 30
4.4. Integração de OCR 30
4.5. *Microsoft Azure AI Document Intelligence* 32
4.6. Alinhamento entre Manuscritos e Impressos 32
4.6.1 Definição da Proporção entre Manuscritos e Impressos 33
4.6.2 Aplicação de *Data Augmentation* 33
4.6.3 Validação e Ajustes Finais 34
4.7. Tratamento de Erros com *Ground Truth* 34
4.8. Construção dos Modelos Finais 35
4.9. Integração da Identificação do Autor e da Transcrição de Manuscritos 36
CAPíTULO 5 37
Demonstração 37
5.1. Fundamentação da Demonstração 37
5.2. Desafios e Soluções para Apontamentos nas Universidades 37
5.3. Demonstração Prática no Contexto Universitário 38
5.3.1 Primeiro Cenário: Transcrição de Conteúdos do Quadro 39
5.3.2 Segundo Cenário: Partilha de Apontamentos entre Estudantes 40
5.4. Análise de Resultados 41
5.4.1. Resultados na Identificação do Autor 41
5.4.2 Resultados na Transcrição de Manuscritos 41

5.4.3 Conclusões dos Testes nos Cenários de Demonstração 42
CAPÍTULO 6 43
Avaliação 43
6.1. Avaliação de Requisitos 44
6.2. Limitações 46
CAPÍTULO 7 47
Conclusões 47
7.1. Conclusão 47
7.2. Trabalho Futuro 48
Referências Bibliográficas 49

Índice de Figuras

Figura 1 - Framework da identificação do autor [22], adaptado pelo autor 13
Figura 2 - Fluxograma PRISMA 15
Figura 3- A arquitetura geral do MANCAPT, criado pelo autor 21
Figura 4 - A arquitetura do InceptionV3 [26] 22
Figura 5 – A arquitetura do Azure Document Intelligence para modelos personalizados [27], adaptado pelo autor 23
Figura 6 - Pipeline do protótipo MANCAPT 25
Figura 7 - Amostra de caligrafia escrita em ritmo normal 26
Figura 8 - Linha recortada e tratada de uma amostra de caligrafia 27
Figura 9 - Amostra de uma frase manuscrita, escrita por diferentes autores e o texto impresso correspondente 32
Figura 10 - Teste de reconhecimento de caracteres após alinhamento de manuscritos e impressos . 34
Figura 11 - Correção manual de amostra com resultado de reconhecimento errado 35

Índice de Tabelas

Tabela 1 - Termos de pesquisa e resultados das bases de dados 14
Tabela 2- Tópicos encontrados na síntese crítica da literatura 15
Tabela 3- Lista de requisitos 20
Tabela 4 - Fases de Implementação do Processo de Identificação do Autor 28
Tabela 5- Hiperparâmetros da Identificação do Autor 29
Tabela 6 - Precisão do treino e validação dos modelos VGG16, VGG19 e InceptionV3 30
Tabela 7- Fases de implementação do OCR e Integração com Identificação do autor 31
Tabela 8 - Aplicação de data augmentation sobre amostras manuscritas e impressas 33
Tabela 9 - Resultados da identificação do autor 41
Tabela 10 - Resultados modelo genérico Azure AI Document Intelligence VS modelos MANCAPT 42
Tabela 11 - Feedback dos utilizadores, comparação MANCAPT VS Modelo Geral Document Intelligence 44
Tabela 12 - Estado de execução dos requisitos 45

Lista Abreviaturas

AI	Inteligência Artificial
API	Application Programming Interface
CER	Character Error Rate
CNN	Convolutional Neural Network
DSRM	Design Science Research Methodology
GPU	Graphics Processing Unit
HMM	Hidden Markov Models
LSTM	Long Short-Term Memory
OCR	Optical Character Recognition
RNN	Recurrent Neural Network
SVM	Support Vector Machines

CAPÍTULO 1

Introdução

1.1. Motivação

Apesar dos avanços na digitalização e gestão eletrónica de documentos, a presença substancial de documentação manuscrita em diversos sectores continua a ser um desafio significativo. Sectores como a saúde, o direito, a educação e a administração pública dependem da precisão na conversão desses documentos para formatos digitais. Uma das principais dificuldades enfrentadas nos sistemas de reconhecimento ótico de caracteres (OCR) existentes é a variabilidade da caligrafia individual, que pode variar não apenas de indivíduo para indivíduo, mas também dentro dos escritos de uma mesma pessoa, dependendo de fatores como a pressa ou o estado emocional [1].

Esta variabilidade resulta em altas taxas de erro de reconhecimento, levando a interpretações incorretas de informações críticas, com potenciais consequências graves em processos importantes [2].

Além dos desafios de transcrição, enfrentamos também o problema da gestão de múltiplos documentos, num volume significativo de dados ainda permanece em formato manuscrito [3], representando um vasto arquivo de informações valiosas que ainda não completaram a transição para o digital. Este acervo abrange desde registos médicos em hospitais que datam de décadas, até arquivos jurídicos e contratos importantes armazenados em formatos que não são apenas difíceis de aceder, mas também suscetíveis a danos e perdas ao longo do tempo [4].

O protótipo MANCAPT pretende responder a estes desafios com uma tecnologia de OCR avançada que se adapta à variabilidade da caligrafia, minimiza erros de transcrição e garante ainda a autenticidade dos documentos ao identificar a autoria de manuscritos.

1.2. Contexto

1.2.1 Transição do Analógico para o Digital

Vivemos numa Era de transformação digital, na qual a capacidade de aceder, analisar e partilhar informações de forma rápida e eficaz é fundamental para a competitividade empresarial. Apesar dessa crescente digitalização, uma quantidade significativa de documentos importantes continua retida em formatos analógicos. Este número de documentos não digitalizados não representa apenas um desafio logístico, mas também uma barreira significativa para organizações que procuram eficiência e agilidade nos seus processos [5].

A transição de documentos em papel para formatos digitais é agora uma prioridade global, impulsionada pela necessidade de melhorar o acesso à informação e por exigências regulatórias cada vez mais rigorosas em termos de gestão e armazenamento de dados [6]. Contudo, muitas organizações deparam-se com dificuldades nesta transformação devido ao volume massivo de documentos históricos, registos legais, informações médicas e outros dados críticos que ainda necessitam de ser convertidos em formatos digitais acessíveis e seguros [7].

Sistemas avançados de OCR são essenciais para organizações que gerem grandes volumes de documentos. Permitem a digitalização de dados críticos, preservando a integridade e acessibilidade das informações. "A implementação de tecnologia OCR não só facilita a conformidade com normas regulatórias, mas também promove uma maior eficiência operacional ao transformar documentos analógicos em formatos digitais de fácil acesso", afirma [6] no seu estudo sobre os fatores que impactam o sucesso de projetos de digitalização.

1.2.2 Importância da Escrita Manual na Era Digital

O estudo [8] realça que a fluência da escrita à mão e a precisão ortográfica são essenciais, influenciam diretamente a qualidade da escrita e facilitam processos cognitivos avançados como o planeamento e a tradução de ideias.

Isto desafia a noção de que a escrita manual está obsoleta, esta habilidade demonstra ser crucial para o desenvolvimento educacional e cognitivo, além facilitar o reconhecimento e a memória visual a longo prazo, essenciais na Era Digital onde a informação visual é predominante [9].

No ambiente empresarial, onde a eficiência e a rapidez são valorizadas, a escrita manual ainda é uma ferramenta poderosa devido à sua flexibilidade e adaptabilidade. Para [10] os documentos manuscritos, como notas e *post-its*, são comuns porque permitem a rápida anotação e partilha de informações essenciais sem a rigidez das plataformas digitais.

As empresas frequentemente enfrentam a tarefa de integrar conhecimento não estruturado em seus processos formais. A escrita manual oferece uma maneira de documentar rapidamente ideias e variações nos processos que não se encaixam facilmente nos sistemas digitais estruturados. A possibilidade de incorporar anotações manuais diretamente nos sistemas de gestão de processos empresariais é discutida como uma forma de capturar a rica camada de conhecimentos informais que podem ser importantes para a operação e inovação nos negócios.

Um dos desafios mais significativos na gestão empresarial moderna é a convergência de informações manuais e digitais. Para [10] a interação baseada em papel e caneta ainda é relevante no design de sistemas de negócios, servindo como um meio flexível e acessível comparado aos sistemas digitais, que muitas vezes não podem replicar a simplicidade e eficácia da escrita manual para certas tarefas.

1.2.3 Importância da Identificação do Autor

Uma das maiores dificuldades no campo da Identificação do Autor, apesar das técnicas tradicionais distinguirem diferentes autores, prende-se com a dificuldade em lidar com a variabilidade de escrita do mesmo autor, devido à falta de análise idiossincrática. O mesmo autor, pode variar a sua caligrafia consoante fatores mecânicos, físicos e psicológicos, afetando a precisão dos sistemas de identificação [11]. Além disso, a semelhança entre estilos de escrita pode comprometer a eficácia dos sistemas tradicionais de identificação, tornando difícil distinguir entre diferentes escritores. Outro obstáculo relevante é o impacto de ruídos introduzidos durante a digitalização, que podem distorcer padrões manuscritos e afetar a precisão dos modelos utilizados para identificação. Esses desafios demonstram que os métodos convencionais de identificação de autores não são suficientes para a correta atribuição de autoria, ignoram as características idiossincráticas da escrita manual.

A adoção de *Convolutional Neural Networks* (CNNs) tem demonstrado melhorias na identificação de autores, uma vez que permite capturar traços estilísticos únicos, diferenciar escritores mesmo em pequenas amostras e analisar padrões idiossincráticos de forma automatizada. Medidas como espaçamento entre palavras e dimensões dos caracteres mostraram-se eficazes na diferenciação de escritores, reduzindo a dependência de longas amostras de texto e aumentando a capacidade dos modelos de identificação [12].

1.3. Objetivos

Tendo em conta o contexto e a motivação anteriormente descritos, este trabalho consiste no desenvolvimento do protótipo MANCAPT é motivado pela necessidade de melhorar a eficiência, a precisão e na gestão de documentos manuscritos, áreas que ainda enfrentam desafios significativos apesar dos avanços na tecnologia de digitalização e reconhecimento de texto.

Este trabalho visa desenvolver o protótipo MANCAPT, que tem os seguintes objetivos:

- **Objetivo 1:** Identificar o autor dos documentos manuscritos através da caligrafia do texto;

- **Objetivo 2:** Transcrever textos manuscritos por pessoas para texto digital, tendo em conta a diversidade de letras e estilos pessoais.

Este trabalho representa um marco importante na gestão de documentos manuscritos, promovendo uma transição eficiente para o digital. Com capacidades avançadas de transcrição e identificação de autores, a ferramenta melhora a precisão documental, essencial em várias indústrias onde ainda existe dependência da escrita manual ou da sua transcrição para dados digitais.

1.4. Metodologia

O protótipo MANCAPT foi desenvolvido com base na metodologia *Design Science Research Methodology* (DSRM), esta é utilizada para criar e avaliar artefactos que abordam problemas específicos, combinando teoria e prática. A DSRM é particularmente eficaz para a criação de artefactos digitais que exigem soluções inovadoras e empíricas, oferecendo uma abordagem estruturada para a pesquisa e desenvolvimento [13].

A DSRM é composta por seis fases principais, cada desempenhando um papel relevante no processo de pesquisas:

1. **Identificação e motivação do problema:** O problema identificado consiste nas limitações dos sistemas de OCR em lidar com a variabilidade da caligrafia individual;
2. **Definição dos objetivos da solução:** O principal objetivo é desenvolver uma solução capaz identificar automaticamente o autor e transcrever manuscritos com elevada precisão.;
3. ***Design* e desenvolvimento:** É concebido o protótipo MANCAPT com dois componentes principais: um modelo baseado em *InceptionV3* para identificação autoral, e modelos personalizados com o *Microsoft Azure Document Intelligence* para a transcrição. O desenvolvimento envolve recolha e tratamento de dados, técnicas de pré-processamento (binarização, remoção de ruído, normalização) e *data augmentation*.;
4. **Demonstração:** O sistema é testado em dois cenários em contexto universitário - transcrição de conteúdos escritos no quadro e partilha de apontamentos manuscritos entre estudantes - com o objetivo de validar a aplicabilidade do protótipo em contextos reais;

5. **Avaliação:** A eficácia do sistema é avaliada com base na capacidade de identificar corretamente os autores dos manuscritos e melhorar a precisão da transcrição face a soluções genéricas. A avaliação contempla comparações com o modelo genérico do *Microsoft Document Intelligence*, bem como a recolha de *feedback* dos utilizadores sobre a usabilidade e aplicabilidade prática do sistema;
6. **Comunicação:** Os resultados do trabalho desenvolvido estão documentados e serão posteriormente comunicados à comunidade académica, com vista à sua publicação e discussão. Adicionalmente, será disponibilizado o MANCAPT dataset, constituído por amostras de caligrafia manuscrita recolhidas no âmbito do MANCAPT, contribuindo para o avanço da investigação na área do reconhecimento ótico de caracteres.

1.5. Estrutura do Documento

Este documento está organizado em sete capítulos, cada um abordando aspetos do desenvolvimento do protótipo MANCAPT. O Capítulo 1 apresenta o contexto, o problema, a motivação, os objetivos e a metodologia. Discutem-se desafios da digitalização de manuscritos e a importância da identificação autoral nos sistemas OCR. O Capítulo 2 é dedicado à revisão da literatura e aborda conceitos e metodologias do OCR, explorando técnicas de pré-processamento, modelos de aprendizagem profunda e identificação de autores. O Capítulo 3 - Desenvolvimento detalha a construção do protótipo MANCAPT, incluindo requisitos, recolha e tratamento de dados, e conceção da arquitetura. Apresentam-se desafios e soluções para otimização do sistema. O Capítulo 4 - Implementação descreve o treino dos modelos, a integração da identificação de autor e a transcrição de manuscritos, abordando alinhamento, tratamento de erros e ajustes finais. O Capítulo 5 - Demonstração valida o sistema em cenários reais, demonstrando a eficácia do protótipo MANCAPT na identificação autoral e transcrição, com análise de resultados e comparação com OCR genéricos. O Capítulo 6 - Avaliação analisa o desempenho do sistema, discutindo métricas como precisão na identificação do autor e taxa de erro na transcrição, além de limitações e melhorias futuras. Por fim, o Capítulo 7 - Conclusões sintetiza os contributos do trabalho, destacando avanços na personalização do OCR para manuscritos e perspetivas de evolução do protótipo MANCAPT. O documento termina com as Referências Bibliográficas, onde são listadas todas as fontes consultadas ao longo do trabalho, garantindo o rigor académico da investigação.

CAPÍTULO 2

Estado da Arte

2.1. Introdução ao Reconhecimento Ótico de Caracteres (OCR)

O OCR é uma tecnologia para a conversão de textos manuscritos e impressos para formato digital. Essa transformação é essencial em diversos setores, como a digitalização de arquivos históricos, automação de processos administrativos e sistemas de leitura de documentos em contextos jurídicos, médicos e educacionais [3, 14].

A tecnologia OCR utiliza técnicas avançadas de processamento de imagem e reconhecimento de padrões para converter imagens digitalizadas em texto editável [15].

Desde os primeiros sistemas OCR, que se baseavam em correspondência de padrões, até aos sistemas modernos que utilizam *deep learning*, a precisão e a eficiência do OCR melhoraram significativamente. Os primeiros sistemas de OCR tinham limitações significativas, especialmente em lidar com variações na caligrafia e a qualidade das imagens [2]. Atualmente, os sistemas OCR modernos utilizam uma combinação de técnicas de *machine learning*, que incluem Redes Neurais Convolucionais (CNN) e Redes Neurais Recorrentes (RNN), para lidar com a variabilidade da caligrafia e a complexidade dos documentos manuscritos [16, 17].

A aplicação do OCR tem se mostrado eficaz em diferentes contextos. No setor médico, por exemplo, os OCR são utilizados para digitalizar registos de pacientes e prescrições, melhorando a acessibilidade e a segurança dos dados [2]. No setor jurídico, facilita a digitalização de contratos e documentos legais, permitindo uma recuperação mais rápida e eficiente das informações [4]. Além disso, no setor educacional, o OCR é utilizado na correção automática de provas e na digitalização de materiais didáticos manuscritos [3].

Vários estudos recentes têm demonstrado que a integração de técnicas de *deep learning*, como CNN e RNN, podem melhorar significativamente a precisão do OCR. As CNN são eficazes na extração de características visuais dos caracteres manuscritos [15], enquanto as RNN são adequadas para capturar dependências sequenciais no texto [16]. Além disso, a combinação dessas técnicas com métodos de pré-processamento de imagem, como remoção de ruído, binarização e ajuste de inclinação, resultam em sistemas OCR mais precisos [14, 18].

2.2. Tecnologias Utilizadas com OCR

2.2.1 Redes Neurais Convolucionais (CNN)

As CNN são amplamente utilizadas para a extração de características visuais dos caracteres manuscritos. Elas são compostas por múltiplas camadas de neurónios artificiais, onde cada camada é responsável por extrair características específicas da imagem, como bordas, texturas e formas [15]. Esta abordagem permite que o modelo aprenda representações hierárquicas das características visuais, melhorando a precisão do reconhecimento.

As CNN operam utilizando três tipos principais de camadas: convolução, *pooling* e *fully connected*. A camada de convolução aplica filtros (*kernels*) para produzir mapas de características, destacando padrões específicos na imagem de entrada. A camada de *pooling* reduz a dimensionalidade dos mapas de características, mantendo as informações mais relevantes e reduzindo a carga computacional. Finalmente, as camadas *fully connected* são responsáveis pela decisão final de classificação, combinando as características extraídas para produzir o resultado de reconhecimento [16].

Os estudos de [16, 5] demonstraram que as CNN são eficazes na extração de características visuais complexas, permitindo um reconhecimento preciso de caracteres manuscritos variáveis. O estudo de [17] utilizou CNN para a deteção automática de assinaturas em documentos, mostrando a versatilidade desta tecnologia.

2.2.2 Redes Neurais Recorrentes (RNN) e Long Short-Term Memory (LSTM)

As RNN são projetadas para processar sequências de dados, tornando-as ideais para o reconhecimento de texto onde a ordem dos caracteres é importante. As RNN mantêm uma memória de informações processadas anteriormente, o que é necessário para capturar dependências de longo prazo no texto manuscrito [2]. No entanto, as RNN tradicionais enfrentam desafios com o desaparecimento do gradiente, dificultando a captura de dependências de longo alcance [3].

Para superar essas limitações, foram desenvolvidas as redes *Long Short-Term Memory* (LSTM). As LSTM são uma variante das RNN que utilizam unidades de memória especializadas para manter informações por longos períodos, evitando problemas de desaparecimento e explosão do gradiente. Esta característica torna as LSTM particularmente eficazes no processamento de texto manuscrito, onde a sequência e o contexto são fundamentais para a compreensão correta dos caracteres [16].

No estudo de [2] foi utilizada uma arquitetura baseada em *VGG16* combinada com Bi-LSTM para o reconhecimento de prescrições médicas manuscritas. A combinação de CNN para extração de características visuais e LSTM para modelação de dependências sequenciais permitiu alcançar níveis de precisão elevadas. O estudo de [19] também explorou o uso de LSTM para capturar nuances na caligrafia, destacando a eficácia desta abordagem para a identificação do autor através da caligrafia.

2.2.3 Modelos Ocultos de Markov (HMM) e Máquinas de Vetores de Suporte (SVM)

Os Modelos Ocultos de Markov (HMM) e as Máquinas de Vetores de Suporte (SVM) são técnicas tradicionais que também são utilizadas no OCR. Os HMM são particularmente eficazes para modelar dependências temporais em sequências de dados, enquanto as SVM são empregues para a classificação de caracteres com base em características extraídas das imagens [18].

Os HMM são uma técnica estatística usada para modelar processos que têm uma estrutura temporal, sendo particularmente útil para o reconhecimento de padrões sequenciais. No contexto do OCR, os HMM são utilizados para reconhecer palavras e frases inteiras ao invés de apenas caracteres individuais. Este método baseia-se na probabilidade de transição entre estados, que pode representar a transição entre diferentes caracteres ou palavras num texto.

A força dos HMM reside na sua capacidade de lidar com variabilidade e incerteza, tornando-os adequados para tarefas onde há dependência temporal e variação natural, como a escrita à mão. Os HMM são eficazes em situações onde o reconhecimento de padrões requer a consideração do contexto anterior e seguinte, permitindo assim uma compreensão mais abrangente do texto [14].

As *Support Vector Machine* (SVM) são uma poderosa técnica de aprendizagem supervisionada utilizada para classificação e regressão. No contexto do OCR, as SVM são empregues para a classificação de caracteres, onde a tarefa é identificar a qual classe (ou letra) um determinado conjunto de características pertence. As SVM funcionam através da construção de hiperplanos no espaço de alta dimensão que separam os diferentes tipos de caracteres. Um dos principais benefícios das SVM é a sua eficácia em espaços de alta dimensionalidade e a capacidade de utilizar funções *kernel* para lidar com a não-linearidade dos dados. As SVM têm mostrado excelentes resultados na classificação de imagens de caracteres devido à sua capacidade de maximizar a margem de separação entre classes, reduzindo assim o erro de generalização [20].

Em alguns sistemas de OCR, uma abordagem híbrida que combina HMM e SVM pode ser utilizada para tirar proveito dos pontos fortes de ambas as técnicas. Os HMM podem ser utilizados para modelar a estrutura sequencial do texto, enquanto as SVM podem ser usadas para a classificação precisa dos caracteres individuais. Esta combinação permite uma maior flexibilidade e precisão, especialmente em sistemas onde a variabilidade na escrita e o contexto são fatores importantes a considerar [18].

2.2.4 Técnicas de Pré-processamento

As técnicas de pré-processamento de imagem são fundamentais para melhorar a precisão dos sistemas OCR. Este processo envolve várias etapas, cada uma focada em preparar a imagem para a extração de características e reconhecimento de texto:

- **Binarização -** A binarização é uma das primeiras etapas no pré-processamento de imagens e envolve a conversão de uma imagem em escala de cinza para uma imagem binária. Este processo é efetuado para facilitar a distinção entre o texto e o fundo da imagem. Métodos comuns de binarização incluem algoritmos de limiarização automática que ajustam o valor de limiar para segmentar a imagem [18];
- **Remoção de Ruído -** A remoção de ruído serve para melhorar a qualidade da imagem antes do reconhecimento. Ruídos podem ser causados por imperfeições no papel, manchas ou defeitos no processo de digitalização. Técnicas comuns de remoção de ruído incluem filtros de média, filtros de mediana e transformadas morfológicas. Estas técnicas ajudam a suavizar a imagem e a remover artefactos indesejados, facilitando o trabalho dos algoritmos de OCR [18];
- **Normalização de Escala -** A normalização de escala ajusta a imagem para um tamanho padrão, garantindo que o texto dentro da imagem tenha uma altura consistente. Isto é importante, porque variações de escala podem afetar negativamente o desempenho dos algoritmos de reconhecimento. Métodos de normalização de escala geralmente envolvem a transformação linear das dimensões da imagem para se ajustar a uma escala predefinida [16];
- **Correção de Inclinação -** A correção de inclinação aborda o problema de imagens de texto que não estão alinhadas corretamente. Textos inclinados podem ser difíceis de reconhecer corretamente, e a correção de inclinação envolve a deteção e ajuste do ângulo da linha base do texto para horizontal. Algoritmos como o *"Hough Transform"* são frequentemente utilizados para detetar e corrigir a inclinação [16];

- **Segmentação -** A segmentação é o processo de dividir a imagem de texto em componentes menores, como linhas, palavras e caracteres. Este passo permite que os algoritmos de OCR processem cada componente individualmente. Métodos de segmentação podem variar desde técnicas baseadas em histogramas até métodos mais avançados que utilizam redes neurais para identificar os limites entre os diferentes componentes do texto [3];
- ***Enhancement* -** O *enhancement* de imagem envolve a aplicação de técnicas para aumentar a qualidade visual da imagem. Isto pode incluir o ajuste de contraste, afiação de bordas e realce de detalhes. Estas técnicas ajudam a tornar o texto mais distinguível do fundo, facilitando o reconhecimento pelos algoritmos de OCR. Estudos indicam que o *enhancement* eficaz pode melhorar significativamente a taxa de sucesso do reconhecimento de caracteres [21].

2.3. Identificação do Autor

A identificação do autor através da caligrafia é um campo de estudo significativo que tem implicações importantes para a análise forense, autenticação de documentos e pesquisa histórica. Técnicas de reconhecimento de padrões e análise de características específicas da escrita são utilizadas para diferenciar e identificar autores com base nas suas características caligráficas únicas [22].

2.3.1 Métodos Baseados em CNN e RNN

As CNN são utilizadas para extrair características visuais detalhadas da caligrafia, enquanto as RNN, especialmente as LSTM, são usadas para capturar dependências sequenciais na escrita. A combinação dessas técnicas permite uma análise detalhada e precisa da caligrafia, facilitando a identificação do autor [22].

No estudo de [22] ficou demonstrado que a utilização de CNN para a extração de características visuais, seguida da aplicação de RNN para analisar a sequência das letras, pode resultar numa identificação autoral precisa. As CNN identificam padrões visuais complexos, enquanto as RNN avaliam como esses padrões se conectam ao longo do tempo, criando autenticidade da caligrafia do autor.

2.3.2 Análise Idiossincrática

A análise idiossincrática foca-se nas particularidades individuais da escrita de cada autor. Estas, podem incluir a forma, tamanho e inclinação das letras, bem como padrões de espaçamento e conectividade entre caracteres. Esta abordagem tem-se mostrado eficaz na diferenciação de

autores em manuscritos, pois a análise dessas características específicas permite a distinção entre diferentes autores com base nas suas idiossincrasias caligráficas [19].

Os autores destacam a importância de características como a pressão aplicada na escrita, fluidez dos traços e a regularidade dos espaços entre palavras e letras. Essas características são únicas para cada indivíduo e, quando analisadas em conjunto, podem oferecer níveis de precisão elevados na autenticidade da escrita.

2.3.3 Métodos de Identificação de Autores

Conforme descrito, em 2017 por [23], os métodos de identificação de autores evoluíram significativamente ao longo dos anos, abrangendo desde abordagens baseadas em perfis até métodos híbridos que combinam várias técnicas. Esta revisão destaca a utilização de conjuntos de características que incluem vocabulário, propriedades sintáticas e outras propriedades textuais para a identificação da autoria. Eles categorizaram os métodos de identificação de autores em três abordagens principais:

- **Abordagens Baseadas em Perfis**: Estas abordagens concatenam todos os textos de um autor num único perfil e comparam novos textos com este perfil usando medidas de similaridade. Modelos probabilísticos e de compressão são frequentemente utilizados nesta categoria;
- **Abordagens Baseadas em Instâncias**: Estas abordagens tratam cada texto individualmente, utilizando modelos de espaço vetorial e algoritmos de *machine learning* para classificar novos textos com base em textos de treino conhecidos;
- **Abordagens Híbridas**: Estas combinam elementos das abordagens baseadas em perfis e instâncias, procurando tirar vantagem dos pontos fortes de ambas as metodologias.

Os autores de [23] sublinham ainda a importância da seleção do conjunto de características e a aplicação de modelos apropriados e técnicas como a análise de n-gramas e modelos de espaço vetorial e para alcançar alta precisão na identificação de autores.

A Figura 1. ilustra um processo típico de identificação de autor baseado na comparação de características extraídas de documentos de treino e um documento de teste. Este processo requer a entrada de documentos de treino, que são textos escritos por diferentes autores. As características do documento de teste (que deve ser diferente do documento de treino), elaborado por um dos autores dos documentos de treino, são extraídas e comparadas com características dos documentos de treino através de computação estatística. Com base nas características de escrita, cada autor recebe uma pontuação, e essas pontuações são utilizadas para verificar a correspondência e identificar o autor mais provável do documento de teste.

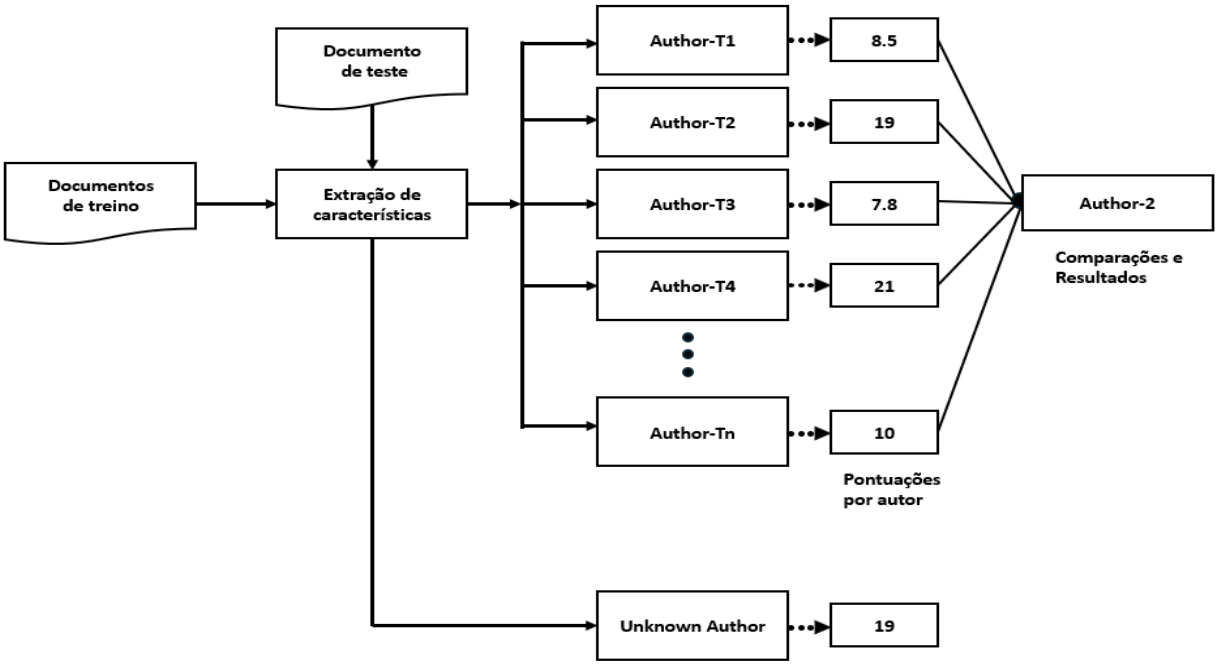


Figura 1 - Framework da identificação do autor [22], adaptado pelo autor

2.4. Síntese Crítica da Literatura

Para análise do estado da arte sobre a tecnologia OCR, utilizou-se a metodologia *Prefered Reporting Items for Systematic Reviews and Meta Analysis* (PRISMA). Esta metodologia é amplamente reconhecida por garantir rigor e transparência em revisões sistemáticas [24]. A pesquisa foi efetuada em 2 bases de dados principais: Scopus e IEEE. Os resultados foram registados na ferramenta *Mendeley* e tratados na ferramenta *Microsoft Excel*. As palavras-chave utilizadas foram "OCR", "*handwritten*", "*identification*" e "*text recognition*", como demonstrado na Tabela 1. Foram estabelecidos também critérios de inclusão e exclusão com o objetivo de aproximar a investigação à orientação do MANCAPT onde o objetivo é apresentar uma solução OCR, com destaque para a identificação do autor.

- **Critérios de inclusão:**
  - Publicações entre 2014 e 2024;
  - Relevância direta ao tema "OCR" e identificação do autor.
- **Critérios de exclusão:**
  - Artigos duplicados;
  - Publicações que não abordassem diretamente OCR ou identificação do autor;
  - Publicações fora do intervalo de data estabelecido.

Tabela 1 - Termos de pesquisa e resultados das bases de dados

Keywords	Base de Dados	Resultados
OCR AND handwritten AND identification AND “text recognition”	SCOPUS	24
OCR handwritten identification "text recognition"	IEEE	105
TOTAL		129

- **Fase de triagem:**
    1. **Identificação:** Foram inicialmente identificados 129 artigos das bases de dados Scopus (24) e IEE (105), utilizando as palavra-chave supracitadas para pesquisa direta e também na forma de *string (“*OCR” AND “*handwritten”* AND “*identification”* AND “*text recognition*”);
    2. **Triagem:** Após a remoção de duplicados, foram analisados 100 artigos;
    3. **Elegibilidade:** Desses, 47 artigos foram considerados válidos com base nos critérios de inclusão, depois, numa segunda análise sistemática foram selecionados 38 artigos: 29 com base nos critérios de inclusão e 9 que mencionaram ou tinham relação direta com a pesquisa por “identificação do autor”;
    4. **Inclusão:** Finalmente, 12 artigos foram selecionados para análise detalhada e através de um website foi incluído um artigo considerado relevante para o objetivo desta análise, totalizando assim 13 artigos incluídos.

A Figura 2 ilustra o fluxograma do processo de síntese crítica da literatura, seguindo a metodologia PRISMA [24].

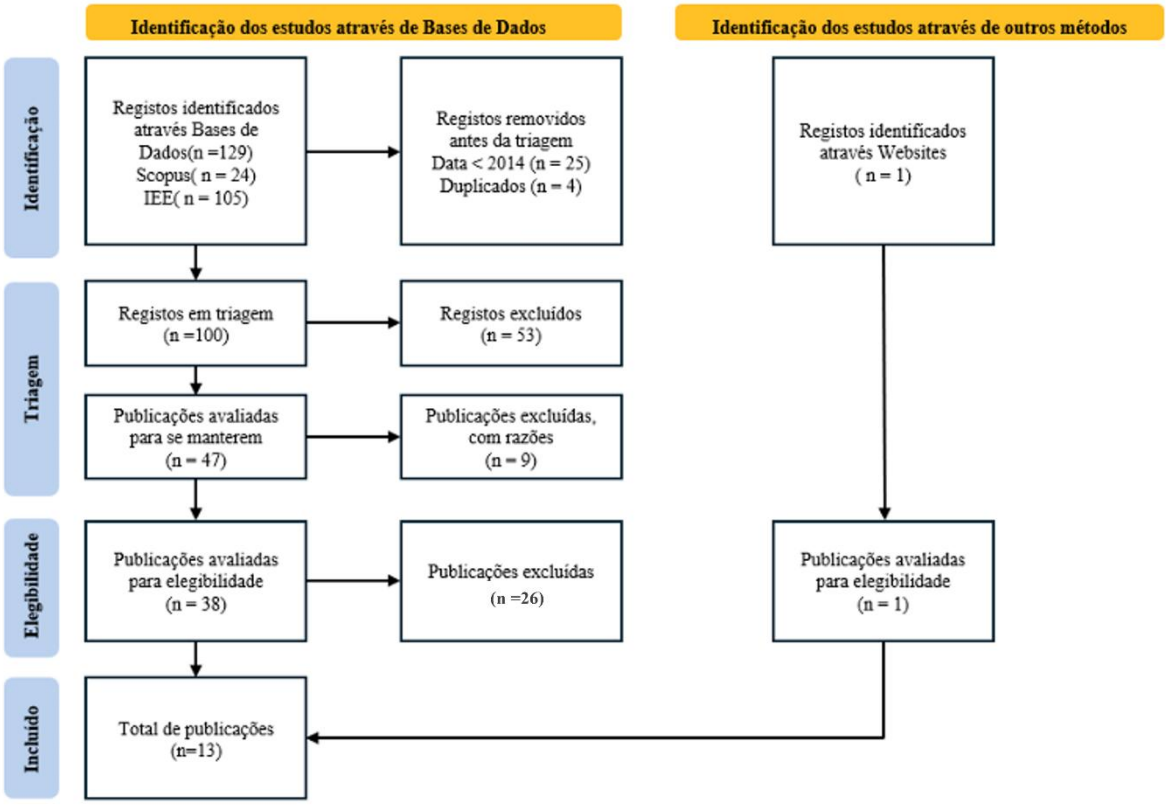


Figura 2 - Fluxograma PRISMA

2.5. Discussão de Resultados

A análise dos artigos selecionados na síntese crítica da literatura foi organizada de acordo com os principais tópicos de pesquisa identificados. A Tabela 2 resume os principais tópicos dos estudos analisados.

Tabela 2- Tópicos encontrados na síntese crítica da literatura

Tópico	Referências	Número de documentos
Reconhecimento de texto	[16, 15, 17, 2, 1, 3, 14,18, 4, 19, 20]	11
Identificação do autor	[21, 1, 2, 22, 4]	5
Técnicas de pré-processamento.	[18, 14, 19, 3, 16, 15, 20]	7
Aplicação multilinguística	[14, 19]	2
Abordagens híbridas	[3, 2]	2
Análise de caligrafia	[1, 22, 21, 16, 17]	5

A maioria dos estudos analisados focou-se no reconhecimento de texto, utilizando diversas tecnologias como CNN e LSTM. Estes estudos demonstraram que as CNN são altamente eficazes na extração de características visuais dos caracteres manuscritos, alcançando precisões superiores a 90%. A aplicação de técnicas de *deep learning* mostrou-se promissora para a melhoria contínua da precisão dos sistemas OCR [15, 16]. Por exemplo, os autores de [15] alcançaram uma precisão de 100% na transcrição de textos em inglês, destacando a eficácia das CNN na extração de características visuais detalhadas. Além disso, no artigo [21] foi apresentada uma abordagem sem segmentação para reconhecimento de frases manuscritas, destacando uma metodologia inovadora que também contribui significativamente para o reconhecimento de texto.

A identificação do autor através da análise caligráfica é outro tópico amplamente explorado. Os artigos [21, 1, 2, 22, 4] destacaram a eficácia da utilização de CNN e RNN para capturar as idiossincrasias da caligrafia individual. Estas técnicas permitiram uma identificação autoral precisa, com estudos relatando precisões que variam entre 83% e 94%. Por exemplo, [19] utilizaram LSTM para identificar nuances na caligrafia, resultando numa precisão de 92%.

Os trabalhos de [21, 14, 19, 3, 16, 15, 20] sublinharam a importância das técnicas de pré-processamento de imagens, como binarização, remoção de ruído e correção de inclinação, para melhorar a precisão dos sistemas OCR. Os autores de [18] demonstraram que a aplicação dessas técnicas pode preparar melhor as imagens para a etapa de reconhecimento, aumentando significativamente a precisão do OCR. A binarização, por exemplo, facilita a distinção entre texto e fundo, melhorando a qualidade da imagem para processamento subsequente.

A aplicabilidade do OCR em diferentes idiomas foi abordada em dois estudos: [14, 19] exploraram a eficácia dos sistemas OCR em contextos multilingues, demonstrando que as técnicas de *deep learning* podem ser adaptadas para lidar com a variabilidade entre idiomas, alcançando precisões elevadas. Por exemplo, [20] alcançaram uma precisão de 98,6% no reconhecimento de caracteres em inglês utilizando SVM, destacando a versatilidade e eficácia dessas técnicas em diferentes contextos linguísticos.

Os estudos de [3, 2] destacaram a utilização de abordagens híbridas que combinam várias técnicas de OCR. Os autores de [3] mostraram que a integração de CNN, RNN e SVM pode resultar em sistemas OCR mais sólidos e precisos. Estas abordagens permitem aproveitar os pontos fortes de cada técnica, oferecendo uma solução mais abrangente e eficaz para o reconhecimento de texto manuscrito. Por exemplo, a combinação de CNN para extração de características e RNN para modelação de dependências sequenciais tem se mostrado particularmente eficaz.

A análise detalhada da caligrafia, com foco nas características individuais da escrita, foi explorada em cinco estudos: [1, 22,21,16, 17] aplicaram técnicas de *machine learning* para identificar padrões únicos na caligrafia, como a pressão da escrita, a fluidez dos traços e a regularidade dos espaços entre letras e palavras. Os estudos de [1, 22] sublinharam a importância de características como a análise de n-gramas e modelos de espaço vetorial para alcançar alta precisão na identificação de autores. Estas características são importantes para distinguir entre diferentes autores, especialmente em contextos forenses e de autenticação de documentos.

Em suma, a síntese crítica da literatura destaca a eficácia e a aplicação diversificada das técnicas de OCR e identificação de autor. A combinação de diferentes métodos tem mostrado resultados promissores, aumentando a precisão e a aplicabilidade do OCR em vários contextos e idiomas. No total, foram analisados 13 artigos na presente revisão.

2.6. Conclusão

A análise detalhada dos artigos selecionados nesta síntese crítica da literatura permite-me concluir que a investigação no campo dos OCR e da identificação de autor através da caligrafia está a avançar de forma significativa. Entre os "*major findings"* desta revisão destacam-se a eficácia e a aplicabilidade das diferentes técnicas e metodologias utilizadas para incrementar a precisão dos sistemas OCR. A pesquisa académico-científica nesta área tem vindo a crescer, refletindo um interesse crescente em melhorar a eficiência dos sistemas OCR, tanto para textos impressos como manuscritos. Os sistemas OCR orientados para a digitalização de manuscritos têm apresentado evoluções notáveis, através da integração de metodologias avançadas de *deep learning* e técnicas de pré-processamento de imagens.

A diversidade de abordagens exploradas nos estudos revistos confirma que não há uma solução única, mas sim um conjunto de estratégias que, quando combinadas, podem superar as limitações individuais e proporcionar resultados mais fiáveis. Esta revisão permite-me afirmar com confiança que o futuro dos sistemas OCR é promissor, com potenciais aplicações amplas e impactantes em vários setores. A continuidade da investigação e desenvolvimento nesta área é essencial para explorar plenamente o potencial desta tecnologia, melhorando ainda mais a sua eficácia e adaptabilidade a novos desafios e requisitos.

CAPÍTULO 3

MANCAPT

Este capítulo introduz e contextualiza as etapas desenvolvidas no âmbito do MANCAPT, nesta fase, o foco esteve na criação do protótipo que serve como solução para o problema identificado. O protótipo desenvolvido consiste num sistema OCR capaz de identificar o autor de um manuscrito através de classificação, selecionar o modelo personalizado correspondente e utilizar o mesmo para transcrever um manuscrito. Esta solução é composta por 2 módulos principais:

- **Módulo de Identificação do Autor** – Responsável por classificar a autoria de um manuscrito utilizando redes neurais convolucionais (CNN), como a arquitetura *InceptionV3*, para extrair padrões característicos da caligrafia e atribuir o manuscrito ao autor correto;
- **Módulo de Transcrição de Manuscritos –** Responsável por selecionar o modelo de transcrição personalizado, correspondente ao autor identificado. Este módulo utiliza o *Azure AI Document Intelligence*, para converter o manuscrito em texto digital, garantindo maior precisão ao adaptar-se ao estilo de escrita específico de cada autor.

3.1. Requisitos

Nesta secção, serão demonstrados na Tabela 3, os requisitos necessários para criar o protótipo MANCAPT, oferecendo uma visão geral das necessidades. Este capítulo visa contextualizar os pressupostos que sustentam o sistema e a definição de critérios que assegurem a eficácia e a usabilidade do sistema. A construção deste protótipo foi guiada pela necessidade de criar uma ferramenta capaz de identificar os autores de manuscritos e transcrever textos manuscritos com elevada precisão.

Tabela 3- Lista de requisitos

ID	Categoria	Descrição
1	Transcrição de Manuscritos	O sistema deve ser capaz de transcrever manuscritos com alta precisão.
2	Identificação do Autor	O sistema deve realizar a identificação do autor de um manuscrito, com base em características únicas.
3	Modelos Personalizados de OCR	Deve haver a possibilidade de criar modelos personalizados de OCR para cada autor identificado.
4	Treino Adaptativo	Treino e adaptação do modelo para cada autor, permitindo a personalização contínua e específica.
5	Integração com *Azure Services*	Integração com o *Azure AI Document Intelligence* para criar um fluxo contínuo de OCR com suporte à personalização.
6	Robustez em Condições Variáveis	O sistema deve ser capaz de lidar com variações de qualidade e estilo dos manuscritos.
7	Suporte a Múltiplos Idiomas	Suporte a múltiplos idiomas para garantir versatilidade em diferentes contextos.
8	Escalabilidade	O sistema deve ser escalável, suportando a inclusão de novos autores de forma ágil.
9	Segurança e Privacidade	Deve garantir segurança na transmissão e armazenamento dos manuscritos digitalizados.
10	Flexibilidade	O sistema deve ser flexível, permitindo ajustes e melhorias ao longo do tempo.
11	Eficiência e Baixa Latência	Baixa latência no processamento, garantindo resultados rápidos e eficientes.

3.2. Desenho da Arquitetura

3.2.1 Arquitetura Geral do MANCAPT

A arquitetura geral do sistema surge como uma resposta integrada, uma abordagem modular que permit**e** uma adaptação dinâmica à variabilidade da escrita de cada autor e aumenta a precisão do reconhecimento textual. A Figura 3, ilustra a estrutura desta arquitetura, detalhando a interação entre os componentes que possibilitam este fluxo de trabalho automatizado e eficiente. A arquitetura do protótipo MANCAPT assenta na integração de dois módulos principais:

- Identificação do autor (*InceptionV3*);
- Transcrição de manuscritos (*Azure AI Document Intelligence*).

A identificação do autor desempenha um papel essencial no processo, ao determinar qual modelo de OCR será usado, o que permite uma transcrição mais precisa e personalizada. Esta abordagem diferencia o MANCAPT de sistemas tradicionais, ao garantir que cada manuscrito seja processado de acordo com as características individuais do autor.

O módulo de identificação do autor utiliza o *InceptionV3* para classificar a autoria do manuscrito com base num conjunto de treino previamente anotado e direciona a transcrição para o modelo personalizado do *Azure AI Document Intelligence*. O módulo de transcrição aplica o modelo específico de OCR associado ao autor identificado, convertendo o manuscrito em texto digital de forma mais precisa e eficiente. Esta arquitetura modular permite não apenas uma transcrição mais fiável, mas também a possibilidade de expandir a base de autores reconhecidos, tornando o sistema escalável.

Figura 3- A arquitetura geral do MANCAPT, criado pelo autor

3.2.2 Identificação do Autor com *InceptionV3*

Inicialmente, foram testados os modelos *VGG16* e *VGG19* para a tarefa de identificação do autor através de classificação. Estes modelos, conhecidos pela sua arquitetura profunda e eficaz na extração de características visuais, foram utilizados como ponto de partida para avaliar a viabilidade da classificação de manuscritos.

No entanto, apresentaram algumas limitações ao lidar com a complexidade dos manuscritos. A rede conseguiu extrair características gerais da escrita, mas não foi suficiente para capturar as nuances específicas da caligrafia dos diferentes autores. Foram feitas várias iterações de treino, utilizando *data augmentation* para criar variações nas imagens de treino, como rotação, brilho, e adição de ruído. Apesar de algumas dificuldades, existiu algum sucesso inicial, mas a precisão não foi suficiente para a tarefa específica de identificação de autor, principalmente devido à falta de capacidade para diferenciar nuances subtis entre escritas semelhantes.

Com os resultados limitados obtidos pelos modelos *VGG*, optou-se por utilizar a rede neural *InceptionV3***.** Esta rede possui uma arquitetura mais avançada, capaz de capturar detalhes complexos através de uma combinação de diferentes convoluções em paralelo[25], tornando-a ideal para a identificação de padrões intricados na caligrafia.

Este modelo contém mais de vinte milhões de parâmetros e destaca-se pela sua arquitetura eficiente na extração de características visuais. Desenvolvido com blocos simétricos e assimétricos, integra múltiplas operações como convoluções, *pooling*, *dropout* e camadas *fully connected*. A Figura 4. apresenta a arquitetura do modelo, evidenciando a organização dos seus blocos e camadas. A *batch normalization* é aplicada para otimizar a aprendizagem, e a classificação final é realizada através da função *Softmax*[26].

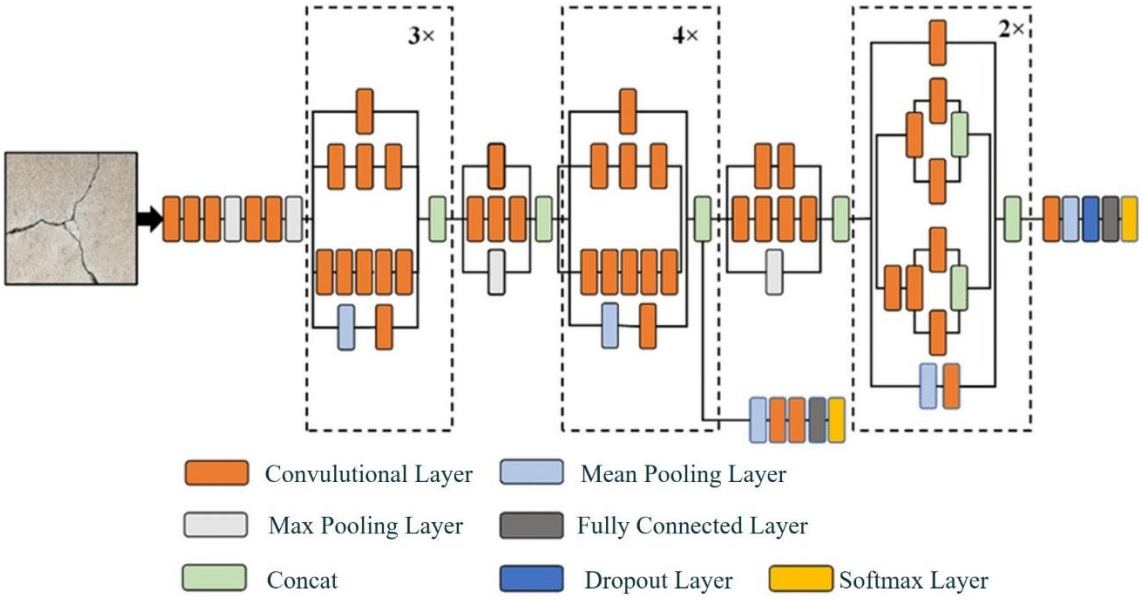


Figura 4 - A arquitetura do InceptionV3 [26]

3.2.3 Transcrição de Manuscritos com *Azure AI Document Intelligence*

Ao longo da criação do protótipo MANCAPT, foi essencial considerar abordagens alternativas antes de consolidar a escolha pelo *Microsoft Azure AI Document Intelligence*. Inicialmente, explorámos modelos baseados em redes neuronais profundas, nomeadamente a *Convolutional Recurrent Neural Network* (CRNN) e *o Transformer Optical Character Recognition* (TROCR), desenvolvido pela *Microsoft*. Estas tecnologias são reconhecidas pela sua elevada capacidade em reconhecer textos manuscritos, especialmente devido ao uso de convoluções para extração de padrões visuais e redes recorrentes ou transformadores para captura da sequência e do contexto textual.

Contudo, ambas as abordagens apresentaram um desafio significativo: a necessidade de um volume massivo de dados anotados para obter resultados consistentes. No contexto do MANCAPT, onde o objetivo é identificar e transcrever caligrafias individuais com alta precisão, a recolha e anotação de dados em larga escala revelaram-se inviáveis em termos de tempo e recursos. Além disso, mesmo utilizando técnicas de *data augmentation*, os modelos demonstraram dificuldades em capturar as nuances da caligrafia dos autores com o conjunto de dados limitado disponível.

O Document Intelligence da *Microsoft Azure* foi escolhido como a plataforma central para operacionalizar a integração entre a identificação do autor e o OCR. A sua capacidade de criar modelos customizados permitiu a personalização do processo para cada autor identificado, ajustando-se às especificidades de cada caligrafia. A integração iniciou-se com o armazenamento das imagens manuscritas e impressas no *Azure Blob Storage*, onde foram organizadas em contentores individuais para cada autor. Este método assegurou um fluxo estruturado, como demonstrado na Figura 5, essencial para o desenvolvimento do protótipo MANCAPT.

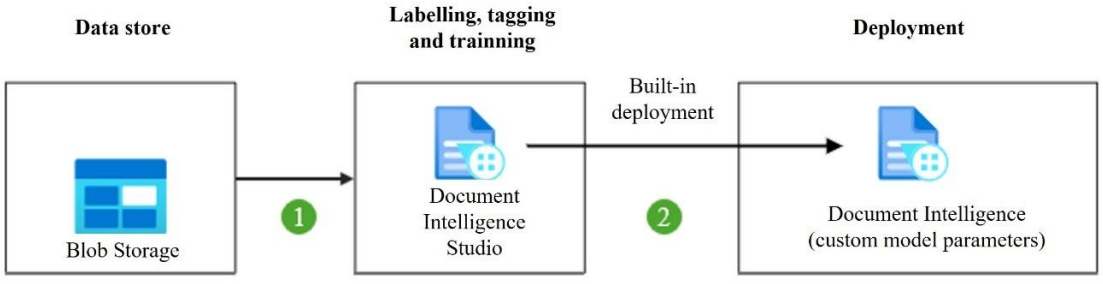


Figura 5 – A arquitetura do Azure Document Intelligence para modelos personalizados [27], adaptado pelo autor

CAPÍTULO 4

Implementação

Este capítulo acompanha o percurso desde a recolha de amostras até à integração final dos modelos, como ilustrado na Figura 6. A criação do MANCAPT dataset, revelou-se essencial para garantir a adaptação do modelo à variabilidade caligráfica. Apesar do volume de dados ser modesto, privilegiou-se a qualidade e adequação ao problema, o que se traduziu em melhores resultados na identificação e transcrição. O capítulo inicia com a recolha de amostras, detalhando o processo de obtenção dos manuscritos e a diversidade caligráfica incluída. Segue-se o tratamento dos dados, com técnicas de pré-processamento como binarização, remoção de ruído e normalização da escala, para melhorar a qualidade das imagens. Posteriormente, descreve-se o processo de identificação do autor, baseado em redes neurais convolucionais e extração de padrões caligráficos. Em paralelo, é explorada a integração do OCR através do *Microsoft Azure AI Document Intelligence*, ferramenta de alto nível para a transcrição dos manuscritos. O alinhamento entre manuscritos e impressos revelou-se uma estratégia central para melhorar o treino, ao garantir melhor correspondência entre os tipos de escrita. A correção manual de erros com base em *ground truth* foi determinante para validar os resultados e aumentar a fiabilidade da transcrição. O capítulo conclui com a construção dos modelos finais e a integração dos módulos de identificação e transcrição, permitindo ganhos significativos na precisão do reconhecimento de manuscritos.

Figura 6 - Pipeline do protótipo MANCAPT

4.1. Recolha de Amostras

Para realizar este trabalho, foram recolhidas amostras de caligrafia de oito pessoas, onde, em seis folhas, escreveram em cada duas um ritmo diferente: baixo, médio e alto. Cada folha continha oito frases, permitindo recolher 16 frases em cada um dos ritmos definidos. Estas frases eram curtas, aleatórias e escritas em português. O objetivo desta recolha foi obter diferentes tipos de letra, considerando o impacto da variabilidade na velocidade e as diferenças causadas na escrita, esperando atingir o maior intervalo de possibilidades e variedade de caligrafia de cada autor.

As imagens como ilustrado na Figura 7, foram disponibilizadas através de dispositivos diferentes, como *smartphones* e impressoras.

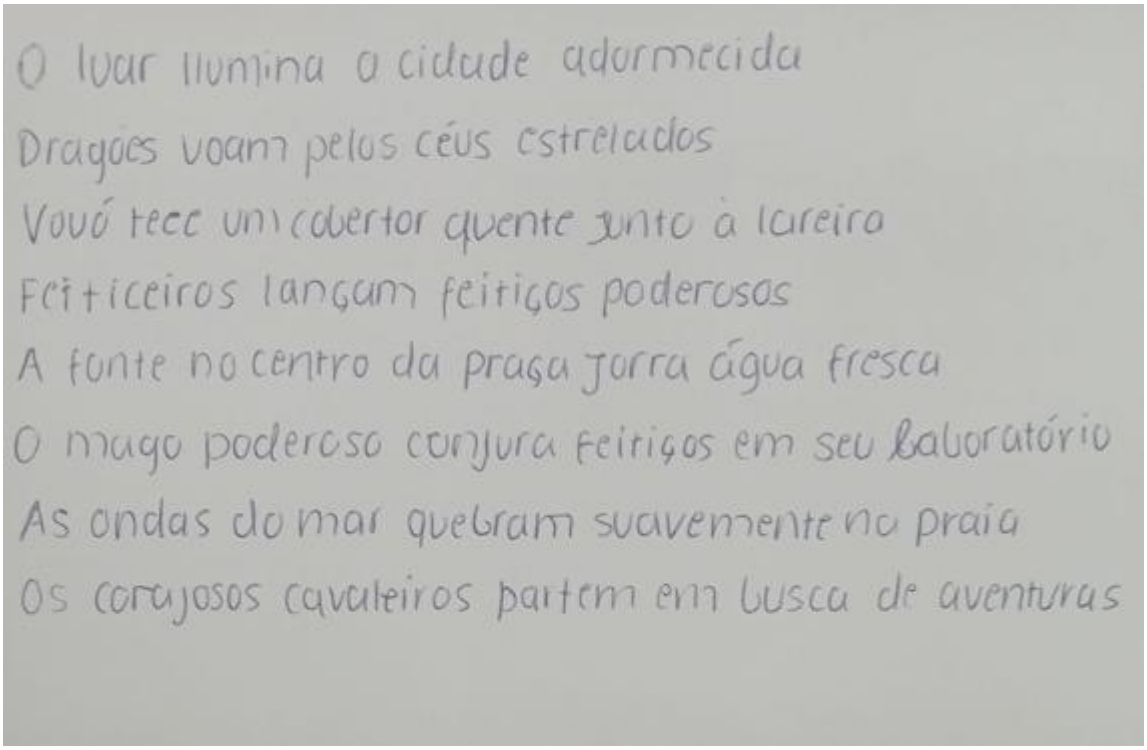


Figura 7 - Amostra de caligrafia escrita em ritmo normal

4.2. Tratamento de Dados

O tratamento dos dados foi importante para garantir que as imagens utilizadas no treino dos modelos estivessem na melhor condição possível. Esse processo envolveu diversas etapas que visaram melhorar a qualidade das imagens e prepará-las para os desafios do reconhecimento e classificação de caligrafia.

Primeiramente, foi realizado um pré-processamento das imagens, que consistiu em cortar e ajustar cada amostra para garantir que apenas a parte relevante da caligrafia fosse mantida. Isto ajudou a minimizar elementos externos que pudessem interferir na aprendizagem do modelo, tais como bordas ou ruídos desnecessários.

Além disso, houve a necessidade de padronizar o tamanho das imagens. Todas as amostras foram redimensionadas para garantir que fossem consistentes em termos de proporções e dimensões. Isso facilitou o processo de treino do modelo, uma vez que entradas uniformes ajudam na eficiência da utilização da memória e aumentam a precisão ao evitar distorções durante a análise.

O ajuste do contraste e do brilho foram também parte do pré-processamento. Este ajuste foi feito de forma manual, utilizando o *MS PowerPoint*, para assegurar que cada frase estivesse clara, com a caligrafia nítida e um fundo branco consistente, como ilustrado na Figura 8.

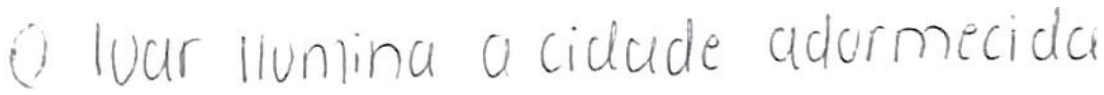


Figura 8 - Linha recortada e tratada de uma amostra de caligrafia

Essa etapa visava tornar os traços da escrita mais destacados, ajudando o modelo a capturar as nuances específicas de cada autor, como a espessura das linhas e o espaçamento entre caracteres.

Por fim, os dados tratados foram organizados e divididos em conjuntos de treino, validação e teste num total de 384 imagens individuais. Esta divisão foi feita de forma cuidadosa para garantir que houvesse uma boa representação das diferentes variações de caligrafia em cada conjunto:

- Conjunto de treino: 230 imagens (60%);
- Conjunto de validação: 58 imagens (15%);
- Conjunto de teste: 96 imagens (25%), não incluídas no processo de treino ou validação.

Esta etapa de tratamento dos dados garantiu uma avaliação eficaz do desempenho do modelo durante o treino e garantindo que os resultados fossem consistentes.

4.3. Processo de Identificação do Autor

Para a implementação do processo de identificação dos autores, foi escolhido o modelo *InceptionV3*, ajustado para classificação de múltiplas classes, cada uma delas representando cada autor. Foram necessárias seis fases, como demonstrado na Tabela 4, para realizar este módulo.

Tabela 4 - Fases de Implementação do Processo de Identificação do Autor

Fase	Ações
Fase 1: Pré-processamento dos dados	Carregamento das imagens e aplicação de aumento de dados para melhorar a consistência do modelo.
Fase 2: Configuração dos geradores de dados, treino e validação	Configuração do gerador dinâmico de aumento de dados. Criação dos conjuntos de treino e validação, garantindo a separação correta dos dados.
Fase 3: Carregamento do modelo base	Utilização do modelo pré-treinado *InceptionV3* para aproveitar características aprendidas.
Fase 4: Adição de camadas personalizadas	Introdução de camadas adicionais para adaptação do modelo ao problema específico.
Fase 5: Compilação do modelo	Definição do otimizador, função de perda e métrica de avaliação para treino eficaz.
Fase 6: Treino do modelo	Treino com validação cruzada e *early stopping* para evitar sobreajuste.
Fase 7: Armazenamento do modelo	Armazenamento do modelo treinado para utilização futura.
Fase 8: Predição e Classificação	Aplicação do modelo para prever a classe da imagem com base nos dados treinados.

4.3.1 Pré-processamento e *Data Augmentation*

As imagens foram organizadas em classes, representando os diferentes autores dos manuscritos, cada amostra foi submetida a um pré-processamento que incluiu o redimensionamento das imagens para 299x299 *pixels*, garantindo a consistência com os requisitos do *InceptionV3*.

As imagens de treino e validação foram sujeitas a *data augmentation* dinâmico em tempo real, recorrendo ao *Keras ImageDataGenerator*, com transformações aleatórias como rotações, deslocamentos, *zoom*, *shear* e espelhamento horizontal. Estas transformações foram aplicadas com o intuito de inflacionar o conjunto de dados [28], no entanto, como é característico do *Keras ImageDataGenerator*, não aumentaram o número de imagens de forma fixa, mas sim geraram variações distintas a cada época de treino.

4.3.2 Treino do Modelo

No treino, o modelo utilizou *categorical crossentropy* como função de perda e *Adam* como otimizador, com uma taxa de aprendizagem ajustada para obter os melhores resultados possíveis. Foi utilizada a técnica de *transfer learning* [29], com *InceptionV3*, partindo de um

modelo pré-treinado no *ImageNet dataset* [30] e adaptando-o para a tarefa de classificação dos autores dos manuscritos.

Este modelo, como os anteriores *VGG16* e *VGG19*, também pode ser considerado pré-treinado, uma vez que utiliza pesos iniciais treinados numa grande base de dados. As camadas finais foram substituídas por camadas específicas para a classificação de oito autores, e o modelo foi treinado para ajustar-se às características únicas de cada um.

4.3.2 Ajuste de Hiperparâmetros

O modelo foi treinado utilizando uma abordagem incremental, onde foram ajustados hiperparâmetros, como demonstrado na Tabela 5, como a taxa de aprendizagem e o número de épocas, de forma a otimizar a precisão na identificação dos autores. Os hiperparâmetros do modelo *InceptionV3* foram selecionados com base em experimentação baseada na literatura existente.

Tabela 5- Hiperparâmetros da Identificação do Autor

Hiperparâmetro	Valor Utilizado	Justificação da Escolha	Referências
Taxa de aprendizagem	0.0001	Permite uma convergência estável e evita oscilações excessivas.	[31]
Tamanho do lote	32	Equilibra consumo computacional e estabilidade do treinamento.	[31]
Número de épocas	20	Balanceia aprendizado adequado e *overfitting*.	[31]
Otimizador	*Adam*	Adaptabilidade dinâmica da taxa de aprendizagem.	[31]
Função de perda	*Categorical Crossentropy*	Padrão para classificação multiclasse.	[25]
Taxa de *dropout*	0.5	Reduz *overfitting* e melhora a generalização do modelo.	[25]
Dimensão das imagens	299×299	Normalização necessária para compatibilidade com *InceptionV3*.	[25]

4.3.4 Resultados do Treino

Foi feita também uma validação contínua, utilizando um conjunto de dados separado, para garantir que o modelo não estava a sobreajustar-se aos dados de treino. Os resultados obtidos com o *InceptionV3*, como demonstrado na Tabela 6, mostraram uma melhoria significativa em relação aos modelos *VGG16* e *VGG19*. A precisão na identificação dos autores aumentou substancialmente, especialmente em casos onde havia grande distinção entre os estilos de escrita. O modelo conseguiu aprender as nuances da escrita de cada autor, incluindo características como a inclinação das letras, o espaçamento entre palavras, e a pressão aplicada na escrita.

Com esta abordagem, que combinou o uso do *InceptionV3* e *data augmentation*, foi possível treinar um modelo, que através do aumento de dados e camadas convolucionais profundas, conseguiu extrair características únicas de cada autor, incluindo padrões de inclinação, espessura dos traços, e espaçamento entre as letras. Isso garantiu uma taxa de reconhecimento elevada e consistente, mesmo quando os autores apresentavam variações subtis na sua escrita. Esta abordagem garantiu que cada autor fosse reconhecido com alta confiança, mesmo quando apresentavam variações subtis na sua escrita.

Tabela 6 - Precisão do treino e validação dos modelos VGG16, VGG19 e InceptionV3

Modelo	Precisão do Treino	Precisão da Validação
VGG16	78,12%	86,67%
VGG19	78,12%	80,00%
InceptionV3	91,20%	96,67%

4.4. Integração de OCR

A integração entre a identificação do autor e o OCR é o alicerce do MANCAPT. O objetivo principal é identificar o autor de um manuscrito e utilizar um modelo personalizado para a transcrição da sua escrita, aproveitando as particularidades estilísticas de cada caligrafia. Essa abordagem não apenas melhora a precisão da transcrição, mas também abre novas possibilidades para a organização e autenticação de documentos manuscritos. Este processo, demonstrado na Tabela 7, teve seis fases de implementação.

Tabela 7- Fases de implementação do OCR e Integração com Identificação do autor

Fase	Ações
1: Configuração do Ambiente no *Azure*	Criação de uma conta no *Microsoft Azure* e ativação dos serviços necessários.
	Configuração do Armazenamento no *Azure Blob Storage* para armazenar imagens de treino.
2: Criação dos Modelos Personalizados	Criação de um modelo personalizado no *Azure AI Document Intelligence* para cada autor identificado.
	Realização de ciclos iterativos de treino, começando com modelos genéricos para estabelecer uma linha de base.
3: Alinhamento entre Manuscritos e Impressos	Implementação de um processo para criar correspondência entre manuscritos e impressos.
	Geração de 1000 imagens impressas e 428 manuscritas com *data augmentation*.
	Manutenção da proporção de 70% impressas e 30% manuscritas para equilibrar o treino.
4: Tratamento de Erros e *Ground Truth*	Criação e revisão do *ground truth*, focando-se na correção das imagens manuscritas.
	Análise manual de 428 imagens manuscritas, com remoção de 40 imagens de qualidade insuficiente.
	Revisão dos ficheiros *JSON* para garantir anotações corretas.
5: Construção dos Modelos Finais	Desenvolvimento de um modelo geral de OCR, utilizando *data augmentation* e *ground truth* corrigido.
	Anotação das imagens problemáticas num ficheiro Excel e remoção das inconsistências do treino.
	Integração do modelo geral com os modelos específicos, para criar os modelos finais, garantindo uma transcrição fiel e precisa.
6: Integração da Identificação do Autor e da Transcrição de Manuscritos	Alinhamento da funcionalidade de identificação do autor com os modelos personalizados gerados no *Azure AI Document Intelligence*.
	Automatização do uso do modelo correspondente sempre que um autor é identificado

4.5. *Microsoft Azure AI Document Intelligence*

Para cada um dos autores, foi criado um modelo personalizado no *Azure AI Document Intelligence*, utilizando as imagens específicas das suas amostras. Durante o processamento, as amostras foram submetidas a uma triagem inicial para garantir consistência e qualidade. Em seguida, passaram por ciclos iterativos de treino. A implementação inicial utilizou modelos genéricos para estabelecer uma linha de base e avaliar a viabilidade do sistema.

Os resultados iniciais indicaram uma confiança geral baixa no desempenho dos modelos, refletindo os desafios de ajustar o sistema às nuances de cada autor. Essa etapa inicial foi essencial para identificar limitações e traçar melhorias futuras na integração e personalização do processo. Essa análise permitiu delinear os próximos passos no desenvolvimento do protótipo MANCAPT, que envolveram a implementação do processo de alinhamento entre manuscritos e impressos, seguido por estratégias específicas para lidar com erros identificados durante a transcrição.

4.6. Alinhamento entre Manuscritos e Impressos

O processo de alinhamento foi desenvolvido com o objetivo de estabelecer uma correspondência rigorosa entre textos manuscritos e suas versões impressas. Este alinhamento tem como objetivo garantir que o modelo de OCR aprenda no treino a reconhecer as características estilísticas de cada autor, baseando-se em referências impressas, como ilustrado na Figura 9. O método adotado foi estruturado considerando a definição da proporção entre os tipos de dados, a aplicação de *data augmentation* sobre o conjunto inicial de treino e validação de amostras manuscritas, e ainda, amostras impressas correspondentes às amostras de caligrafia.

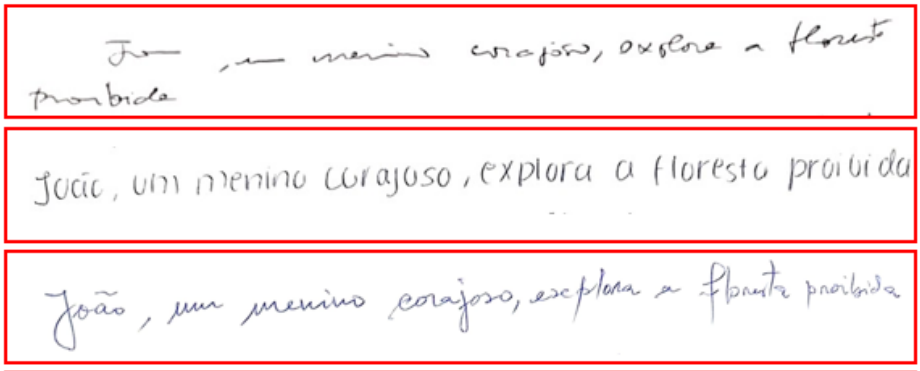
Texto manuscrito Escrito por Diferentes Autores	
Texto Impresso Correspondente	João, um menino corajoso, explora a floresta proibida

Figura 9 - Amostra de uma frase manuscrita, escrita por diferentes autores e o texto impresso correspondente

4.6.1 Definição da Proporção entre Manuscritos e Impressos

Foi escolhida a proporção de 70% de dados impressos e 30% de dados manuscritos com base em iterações experimentais, partindo de um conjunto inicial de 288 imagens manuscritas. Inicialmente, foram testados diferentes rácios, variando entre 50%-50% e 80%-20%. No entanto, verificou-se que a combinação 70%-30% proporcionava um equilíbrio adequado na aprendizagem dos padrões de escrita.

Esse equilíbrio foi alcançado considerando que os textos impressos reforçam a aprendizagem de padrões estruturados e fornecem um referencial para a segmentação e reconhecimento de caracteres, enquanto os textos manuscritos apresentam variações naturais de caligrafia, essenciais para o ajuste fino do modelo.

A definição deste rácio foi validada através da análise da taxa de erro de caracteres (*Character Error Rate* - CER) ao longo dos testes, realizados com amostras do conjunto inicial de testes. Foi observado que abaixo de 30% de manuscritos, o modelo perdia sensibilidade às variações estilísticas, e acima de 30%, a taxa de erro aumentava devido à complexidade dos dados manuscritos.

4.6.2 Aplicação de *Data Augmentation*

Para garantir um conjunto de treino sólido, foi aplicada *data augmentation* tanto às imagens impressas quanto às manuscritas, como demonstrado na Tabela 8. O objetivo foi aumentar a variabilidade dos dados de entrada e expandir o conjunto de treino, simulando diferentes condições de captura e variações naturais de escrita. Foram aplicadas técnicas como ajustes de brilho e contraste para simular diferentes condições de digitalização, rotação e translação para aumentar a generalização do modelo, adição de ruído para refletir variações na qualidade dos documentos digitalizados. Após a aplicação destas técnicas, em 3 iterações, o conjunto de dados final consistiu em 1428 amostras, das quais 1000 eram imagens impressas e 428 manuscritas, geradas com recurso a *data augmentation* a partir do conjunto inicial.

Tabela 8 - Aplicação de data augmentation sobre amostras manuscritas e impressas

Iteração	Manuscritos Iniciais	Impressos Iniciais	Manuscritos Pós Data Augmentation	Impressos Pós Data Augmentation	Total Pós Augmentation	Total de Testes	CER (%)
1	288	0	288	672	960	40	18,2
2	288	672	384	896	1280	50	13,7
3	384	896	428	1000	1428	60	9

4.6.3 Validação e Ajustes Finais

A eficácia do alinhamento de manuscritos e impressos foi validada através de testes de reconhecimento conduzidos no *Azure AI Document Intelligence Studio,* como ilustrado na Figura 10. Além disso, a correspondência entre os textos impressos e manuscritos foi revista manualmente para garantir a integridade dos pares alinhados, assegurando que cada manuscrito possuía uma versão impressa equivalente.

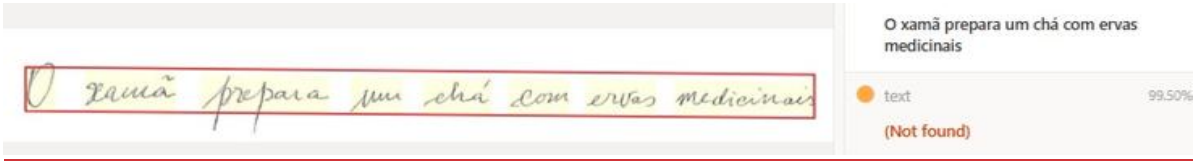


Figura 10 - Teste de reconhecimento de caracteres após alinhamento de manuscritos e impressos

Essa abordagem permitiu influenciar diretamente o conhecimento do modelo sobre os manuscritos, ensinando-o a reconhecer padrões manuscritos ao associá-los às suas equivalências em formato impresso. Assim, o modelo expande a sua compreensão das nuances estilísticas da escrita manual, fortalecendo a capacidade de transcrição com maior fidelidade e consistência.

Dessa forma, garantiu-se que as amostras manuscritas representassem adequadamente a variabilidade da escrita manual no conjunto de treino. O objetivo foi manter uma relação de 70% para dados impressos e 30% para dados manuscritos, garantindo que o modelo tivesse uma base consistente para aprender padrões claros, sem perder a capacidade de generalizar para estilos manuscritos. Esta abordagem foi ajustada iterativamente para maximizar a eficácia do processo de treino e aumentar a capacidade do modelo de lidar com variações estilísticas.

4.7. Tratamento de Erros com *Ground Truth*

No *Azure Document Intelligence Studio,* depois do treino, o sistema executa o reconhecimento sobre todas as amostras e gera uma anotação com o resultado identificado para cada uma, estas anotações ficam armazenadas em arquivos *JSON* e representam a interpretação inicial do modelo e são posteriormente corrigidas no processo do *ground truth* [32], que consiste na correção manual do resultado das anotações geradas com erros após o treino do modelo. Embora o conjunto total incluísse 1428 imagens, o trabalho de correção concentrou-se no subconjunto de 428 imagens manuscritas. Entre essas, foram identificadas cerca de 40 imagens de qualidade insuficiente, que precisaram ser removidas para evitar ruído nos resultados.

Esse processo exigiu uma análise detalhada dos arquivos *JSON* do resultado das anotações manuscritas gerados após o treino do modelo, que nem sempre tinham os resultados desejados, e, portanto, foi necessário corrigir manualmente o resultado de várias anotações para refletirem o reconhecimento pretendido das imagens manuscritas. Para realizar esses ajustes, foi utilizada a ferramenta *Sublime Text*, como ilustrado na Figura 11, o que facilitou a edição e visualização dos dados de forma organizada. Esse trabalho foi particularmente exigente em casos de baixa qualidade das amostras ou ambiguidades.

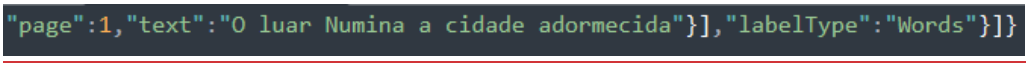


Figura 11 - Correção manual de amostra com resultado de reconhecimento errado.

O processo de treino do modelo envolveu múltiplas iterações de *Ground truth* realizadas diretamente no *Azure Blob Storage*. O *ground truth* serviu para refinar os dados de forma eficiente, assegurando que todas as correções fossem integradas e utilizadas de maneira otimizada pelo *Document Intelligence para melhorar o* treino do modelo.

4.8. Construção dos Modelos Finais

A etapa final do desenvolvimento consistiu na consolidação de um modelo geral de OCR, seguido da integração deste com os modelos personalizados desenvolvidos para cada autor identificado. Este processo visou criar um sistema capaz de adaptar-se às diferentes características das caligrafias manuscritas. O modelo geral, desenvolvido ao longo do processo descrito anteriormente, utilizou técnicas avançadas de *data augmentation* e a aplicação cuidadosa de *ground truth*.

Depois de concluída a criação e validação do modelo geral, cada modelo personalizado foi ajustado seguindo uma metodologia rigorosa. Primeiro, foram identificadas e apontadas num ficheiro *Excel* as imagens que continham erros na transcrição, que foram então excluídas do conjunto de treino. Em seguida, aplicaram-se técnicas de *data augmentation* exclusivamente sobre as imagens corretas para garantir um conjunto de dados confiável e diversificado.

A estratégia final consistiu em combinar no treino, o modelo geral com os modelos personalizados e criar assim os modelos finais dos autores, alinhando-se às exigências do protótipo MANCAPT. Essa dinâmica permitiu que o sistema aprendesse as particularidades caligráficas de cada autor.

4.9. Integração da Identificação do Autor e da Transcrição de Manuscritos

Nesta última fase do processo de desenvolvimento do protótipo, a funcionalidade de identificação do autor foi alinhada com os modelos personalizados, gerados no *Azure AI Document Intelligence* para a transcrição de manuscritos.

Inicialmente, a identificação do autor é feita através do modelo *InceptionV3*, que analisa as características da caligrafia. Após essa identificação, o sistema chama no código, o modelo personalizado correspondente no *Azure AI Document Intelligence*. Esse modelo específico processa a transcrição, garantindo que o reconhecimento do texto seja adaptado ao estilo de escrita do autor identificado.

Dessa forma, ao identificar o autor de um manuscrito, o sistema seleciona o modelo personalizado respetivo, realizando a transcrição do manuscrito com elevada precisão. Esta abordagem permite que o protótipo MANCAPT, cumpra o seu propósito, ao adaptar-se à caligrafia individual de cada autor, para realizar a transcrição de manuscritos.

CAPÍTULO 5

Demonstração

5.1. Fundamentação da Demonstração

A etapa de demonstração é fundamental no contexto da metodologia *Design Science Research Methodology* (DSRM), uma vez que valida o protótipo desenvolvido através de cenários práticos. No MANCAPT, esta fase assume um papel central, dado que permite comprovar a eficácia das soluções propostas para os desafios identificados previamente, nomeadamente:

- A identificação do autor com base na caligrafia;
- A transcrição com elevada precisão de manuscritos.

Neste capítulo, será demonstrado como o protótipo MANCAPT responde a esses desafios, integrando OCR e identificação do autor. Além disso, serão apresentados exemplos concretos de utilização do sistema, acompanhados de uma análise crítica dos resultados.

A demonstração do protótipo MANCAPT será estruturada com base nos seguintes objetivos:

- **Validar o protótipo em cenários reais**: Mostrar como o sistema lida com a transcrição e identificação de autores em situações práticas;
- **Comprovar a eficácia do modelo**: Avaliar o desempenho através de métricas de precisão, tempos de processamento e taxas de erro reduzidas;
- **Identificar limitações e oportunidades de melhoria**: Fornecer uma visão clara dos pontos fortes e áreas que necessitam de melhorias.

5.2. Desafios e Soluções para Apontamentos nas Universidades

Apesar dos avanços tecnológicos na Educação nos últimos anos, nas universidades, os professores ainda recorrem frequentemente à escrita no quadro durante as aulas para transmitir conceitos de forma detalhada. Este método, embora essencial para o ensino, implica desafios significativos devido à rapidez com que os conteúdos são transmitidos. A velocidade necessária para escrever dificulta que os alunos consigam simultaneamente escrever apontamentos e prestar atenção ao que os professores explicam durante as aulas.

Para mitigar estas dificuldades, muitos estudantes optam por fotografar o quadro e posteriormente, utilizam estas imagens para elaborar os seus materiais de estudo. No entanto, interpretar estas fotografias pode ser uma tarefa desafiante, especialmente quando a caligrafia dos professores é menos legível. A velocidade com que os alunos são obrigados a escrever para conseguir acompanhar as explicações orais do professor, faz com que muitas vezes, apenas os próprios consigam perceber os apontamentos que escreveram. Acresce, que estes podem ser partilhados múltiplas vezes por outros colegas, às vezes até de turmas diferentes, sendo assim uma constante a perda da rastreabilidade sobre o autor original dos apontamentos. Este problema dificulta a validação do conteúdo e adicionalmente, impede que se questione o autor original sobre palavras impercetíveis ou abreviaturas utilizadas no momento da escrita, dificultando a compreensão plena dos apontamentos.

Estas dificuldades exigem uma solução que não só melhore a acessibilidade e a legibilidade dos conteúdos manuscritos, mas que também permita identificar claramente o autor dos apontamentos.

O protótipo MANCAPT surge como uma resposta inovadora a estas necessidades. Integrando tecnologias avançadas de OCR, o sistema permite transcrever conteúdos manuscritos com rapidez e precisão. Os professores podem criar modelos personalizados da sua caligrafia, que os estudantes utilizam para converter imagens do quadro em texto digital claro e pesquisável. Esta funcionalidade não só facilita a leitura dos conteúdos como também reduz o tempo necessário para organizar materiais de estudo. Além disso, ao identificar automaticamente o autor dos apontamentos, o sistema reforça a confiança na sua utilização, assegurando que eventuais dúvidas sobre conteúdos ou abreviaturas possam ser dirigidas diretamente ao autor.

5.3. Demonstração Prática no Contexto Universitário

A demonstração prática do protótipo MANCAPT ilustra como o sistema aborda desafios comuns nas aulas universitárias, explorando dois cenários complementares. O primeiro cenário centra-se na transcrição dos conteúdos escritos no quadro pelo professor, permitindo que os alunos registem de forma clara e organizada as informações, mesmo perante a velocidade da escrita do docente. O segundo cenário aborda a partilha de apontamentos entre estudantes, focando-se em situações em que um aluno, ao escrever rapidamente durante as explicações orais do professor, recorre a abreviaturas e notas fragmentadas para acompanhar o ritmo. Esses apontamentos, partilhados posteriormente com outros colegas, enfrentam desafios de compreensão.

Estes cenários refletem como o sistema identifica o autor dos apontamentos e transcreve automaticamente os conteúdos manuscritos. Apesar do elevado nível de fiabilidade das transcrições alcançado pelo protótipo MANCAPT, é importante destacar que o OCR apenas consegue transcrever aquilo que foi efetivamente escrito pelos autores. No entanto, em casos onde algumas palavras são escritas de forma ilegível para o sistema, a capacidade de identificar o autor permite que os utilizadores que realizam transcrição possam identificar e questioná-lo diretamente sobre as palavras impercetíveis.

5.3.1 Primeiro Cenário: Transcrição de Conteúdos do Quadro

Este cenário ilustra como os alunos podem utilizar o MANCAPT para transcrever diretamente o que o professor escreve no quadro durante uma aula. Num ambiente típico de sala de aula, os professores utilizam o quadro para escrever para dar clareza aos conteúdos que estão a transmitir. Essa escrita por norma é feita de forma rápida, obrigando os alunos a acompanhar o ritmo da explicação e a tirar apontamentos simultaneamente. Por isso, muitos alunos recorrem ao telemóvel para tirar fotos do quadro e depois redigir os apontamentos de forma mais pausada. Esta dinâmica pode apresentar alguns desafios à compreensão dos manuscritos devido à velocidade de escrita do professor.

Implementação

1. O professor utiliza o MANCAPT para criar um modelo personalizado da sua caligrafia, baseadas em amostras representativas da sua escrita sugeridas pelo sistema;
2. O sistema, após o processo de tratamento dos dados, adiciona estas amostras ao modelo geral de OCR alinhando os manuscritos aos impressos sugeridos e, por fim, cria o modelo personalizado do professor;
3. Durante a aula, o professor cede aos alunos, o seu ID no MANCAPT, e estes, capturam fotografias dos apontamentos no quadro utilizando dispositivos móveis;
4. Para realizar a transcrição das imagens, os alunos utilizam o ID do professor no MANCAPT, para aceder ao modelo personalizado da sua caligrafia e obter com elevada precisão a transcrição automática dos conteúdos manuscritos nas imagens em texto com alta precisão.

Resultados Esperados

- O sistema alcançou uma taxa de precisão na transcrição de 97%, transformando os apontamentos manuscritos em textos organizados e de fácil leitura;
- Redução significativa de erros na transcrição, especialmente em manuscritos com caracteres rápidos e menos legíveis.

5.3.2 Segundo Cenário: Partilha de Apontamentos entre Estudantes

Neste cenário, o foco recai sobre a partilha de apontamentos entre estudantes, refletindo a dinâmica das aulas em que os alunos precisam de registar rapidamente as explicações orais do professor sem suporte visual no quadro. Esta sobrecarga obriga muitos estudantes a priorizarem uma tarefa em detrimento da outra, gerando apontamentos incompletos ou difíceis de compreender.

Muitas vezes, os alunos dependem de apontamentos feitos por colegas durante as aulas, especialmente porque a velocidade do professor ao falar e a complexidade dos conteúdos forçam os estudantes a registar rapidamente as informações, muitas vezes de forma abreviada ou fragmentada, resultando em apontamentos que apenas o autor consegue compreender completamente. No entanto, a partilha sucessiva desses materiais pode levar à perda de rastreabilidade sobre o autor original, dificultando a validação do conteúdo e a compreensão de palavras ou abreviaturas utilizadas

Implementação

1. Um aluno fez apontamentos das aulas, onde teve de escrever rapidamente, numa escrita não muito clara e com abreviaturas para poder acompanhar a explicação oral do professor;
2. Outros alunos, que não fizeram os apontamentos, pediram para que os cedesse;
3. Depois de acederem decidiram também partilhar com outros colegas;
4. Os alunos que não interagiram com o autor, realizaram a transcrição no MANCAPT;
5. O sistema identificou o autor dos apontamentos manuscritos e realizou a transcrição, mas com algumas palavras impercetíveis e abreviaturas;
6. Os alunos consultaram diretamente o autor para validar e compreender as palavras que lhes levantaram dúvidas.

Resultados Esperados

- A taxa de identificação do autor foi de 95%, permitindo identificar a origem dos apontamentos com elevada precisão;
- Apesar da alta taxa de precisão do OCR (92%) na transcrição, algumas abreviaturas e escrita fragmentada necessitaram da validação do autor para garantir a compreensão total.

5.4. Análise de Resultados

A análise de resultados permitiu comparar os modelos personalizados do protótipo MANCAPT com o modelo genérico do *Azure AI Document Intelligence*, evidenciando o impacto da personalização na melhoria da performance de sistemas OCR. Foram realizados testes com quinze alunos e cinco professores, onde analisámos a precisão da identificação do autor e da transcrição de cada manuscrito, utilizando métricas como a precisão como CER e a taxa de confiança do *Azure AI Document Intelligence*

5.4.1. Resultados na Identificação do Autor

No cenário 2 – partilha de apontamentos entre estudantes, o sistema foi desafiado a reconhecer a autoria de cada manuscrito, atribuindo corretamente o autor a partir do modelo de caligrafia treinado. Os resultados por iteração, como demonstrado na Tabela 9. relevam que quanto maior o número de amostras utilizadas no treino, mais preciso se tornou o modelo.

Tabela 9 - Resultados da identificação do autor

Iteração	Total de amostras	Identificação correta	Identificação incorreta	Precisão
1.ª	10	8	2	80%
2.ª	20	17	3	85%
3.ª	50	49	1	98%

O modelo evoluiu a cada iteração, passando de 80% para 98% com o aumento do número de amostras, melhorando a capacidade do sistema para identificar o autor dos manuscritos e permitindo uma diferenciação mais precisa de manuscritos com caligrafias similares.

5.4.2 Resultados na Transcrição de Manuscritos

No cenário 1 – Transcrição de conteúdos do quadro, o sistema foi testado com os manuscritos dos professores nos quadros captados pelos alunos, com os modelos genéricos do *Azure AI Document Intelligence* e com os modelos personalizados do MANCAPT, verificando-se nos resultados espelhados na Tabela 10, uma maior precisão na transcrição para o formato digital dos modelos personalizados do MANCAPT relativamente ao modelo genérico do *Azure AI Document Intelligence*.

Tabela 10 - Resultados modelo genérico Azure AI Document Intelligence VS modelos MANCAPT

Modelo	Iteração	Amostras	Confiança do Algoritmo (%)	Taxa de Erro de Caracteres (CER)
Document Intelligence Genérico	1ª		89,80%	12,50%
MANCAPT		1008	91,00%	10,20%
Document Intelligence Genérico	2ª		90,10%	11,80%
MANCAPT		1300	93,40%	7,90%
Document Intelligence Genérico	3ª		90,50%	11,30%
MANCAPT		3400	95,80%	6,00%

O modelo proposto neste trabalho superou em todas iterações o modelo genérico do *Azure AI Document Intelligence e* demonstraram que as técnicas utilizadas no MANCAPT como o alinhamento de impressos e manuscritos, *data augmentation* e *ground truth* revelaram-se cada vez mais relevantes para a transcrição, aumentando o número de amostras e garantindo dados "limpos" para melhorar consistência no treino dos modelos personalizados.

5.4.3 Conclusões dos Testes nos Cenários de Demonstração

Os testes realizados nos cenários de demonstração permitiram alcançar as seguintes conclusões:

- O modelo personalizado do MANCAPT superou o modelo genérico do *Azure AI Document Intelligence*, e mostrou-se capaz de reconhecer alguns caracteres pouco legíveis específicos da caligrafia do autor do manuscrito, ao contrário do modelo genérico, apesar de apresentar também um nível de precisão muito alto na transcrição;
- Na transcrição de manuscritos conseguiu reduzir a taxa CER para praticamente metade, de 12,5% para 6% e aumentar a confiança do algoritmo de 89,80% para 95,80% enquanto o modelo genérico manteve-se estável;
- A identificação do autor melhorou a cada iteração passando de 80% para 98% de precisão à medida que mais manuscritos foram incorporados no treino.

CAPÍTULO 6

Avaliação

A avaliação do protótipo MANCAPT constitui uma fase da metodologia DSRM, na qual a avaliação do protótipo desenvolvido serve para validar a sua eficácia. Esta abordagem foi realizada através da recolha de *feedback* de utilizadores que testaram o sistema de OCR personalizado. O objetivo desta análise foi medir a precisão, usabilidade e eficácia do modelo, garantindo que os objetivos definidos fossem atingidos. Para avaliar o desempenho do sistema, foram utilizados dois métodos principais:

- **Testes Práticos** – Os utilizadores transcreveram manuscritos com o MANCAPT e compararam os resultados com o Modelo Geral do *Microsoft AI Document Intelligence*;
- **Inquérito de Avaliação** – Aplicação de um questionário estruturado para recolher perceções sobre a usabilidade, rapidez e precisão do sistema.

A escolha desta metodologia deve-se à necessidade de validar não apenas a precisão técnica do modelo, mas também a experiência de utilização em cenários reais. Assim, o inquérito foi estruturado para obter uma visão quantitativa e qualitativa do desempenho do MANCAPT.

A avaliação contou com a participação dos quinze alunos e cinco professores envolvidos na demonstração, garantindo que a análise refletisse a experiência de um público representativo do uso prático da tecnologia desenvolvida.

O questionário foi dividido em duas secções principais: Identificação do Autor e Transcrição de Manuscritos. As perguntas realizadas foram:

- **Identificação do Autor:**
  - O sistema conseguiu identificar corretamente o autor do manuscrito?
- **Transcrição de Manuscritos:**
  - O sistema conseguiu reconhecer corretamente o conteúdo do manuscrito?
  - Considera que o sistema fornece resultados fiáveis e consistentes?
  - O sistema demonstrou capacidade de reconhecer letras mais complexas, influenciadas pelo estilo de escrita do autor?
  - Recomendaria a utilização deste sistema para transcrição de manuscritos?

Os participantes responderam com base numa escala de *Likert* [33] de 1 a 5, onde 1 representa "Muito insatisfeito" e 5 representa "Muito satisfeito". Os dados recolhidos foram analisados estatisticamente, considerando a média. Foi também realizada uma comparação, referida na Tabela 11, entre o desempenho do MANCAPT e o Modelo Geral do *Document Intelligence.*

Tabela 11 - Feedback dos utilizadores, comparação MANCAPT VS Modelo Geral Document Intelligence

Critério	MANCAPT	Modelo Geral do *Document Intelligence*
Qualidade da Transcrição	4.3	3.9
Precisão do reconhecimento	4.4	3.8
Relevância dos resultados	4.5	3.7
Confiança no sistema	4.2	3.6
Propensão a utilizar novamente	4.6	3.8

Os resultados demonstram que o MANCAPT obteve uma precisão superior ao Modelo Geral do Document Intelligence, apresentando melhor desempenho na transcrição de manuscritos. A análise dos dados recolhidos confirma que os participantes avaliaram positivamente a qualidade da transcrição e a confiabilidade do sistema. De forma geral, a maioria dos participantes recomendaria a utilização do sistema, reforçando a sua aplicabilidade prática.

Os valores detalhados dos resultados encontram-se acima, incluindo as estatísticas de desempenho comparativo. Em suma, a avaliação confirma que os objetivos definidos para o MANCAPT foram alcançados, validando a eficácia do modelo personalizado para reconhecimento de manuscritos.

6.1. Avaliação de Requisitos

A Tabela 12, apresenta a análise do cumprimento dos requisitos do protótipo MANCAPT, considerando os objetivos estabelecidos. Cada requisito foi avaliado com base na sua implementação e eficácia dentro do sistema, classificando-se em Satisfeito, Parcialmente Satisfeito ou Não Satisfeito. Esta avaliação permite identificar áreas onde o sistema já alcançou os resultados esperados e aspetos que ainda necessitam de melhorias ou desenvolvimento adicional.

Tabela 12 - Estado de execução dos requisitos

ID	Categoria	Estado	Descrição
1	Transcrição de Manuscritos	Satisfeito	O sistema realiza transcrição com um alto nível de precisão
2	Identificação do Autor	Satisfeito	A identificação dos autores a partir da caligrafia é realizada com sucesso, permitindo diferenciar entre diferentes escritores.
3	Modelos Personalizados de OCR	Satisfeito	Foram desenvolvidos modelos específicos para diferentes autores, melhorando a precisão da transcrição.
4	Treino Adaptativo	Parcialmente Satisfeito	O sistema permite personalizar o treino por autor, mas ainda não faz ajustes automaticamente com base em novos dados sem intervenção manual.
5	Integração com *Azure Services*	Satisfeito	O sistema está integrado com os serviços *Azure Blob Storage* e *AI Document Intelligence*, aproveitando suas capacidades para armazenamento e processamento de texto manuscrito
6	Robustez em Condições Variáveis	Parcialmente Cumprido	O sistema lida bem com algumas variações de caligrafia, mas ainda pode falhar em casos extremos de ilegível
7	Suporte a Múltiplos Idiomas	Não Satisfeito	Atualmente, o sistema só funciona para um idioma e ainda não suporta múltiplas línguas.
8	Escalabilidade	Parcialmente Satisfeito	O sistema é capaz de lidar com um volume razoável de dados, mas pode enfrentar desafios à medida que o número de documentos cresce, pela necessidade de tratamento manual dos dados
9	Segurança e Privacidade	Parcialmente Satisfeito	O Azure fornece um ambiente seguro, mas não há mecanismos adicionais específicos implementados no MANCAPT.
10	Flexibilidade	Satisfeito	O sistema permite ajustes e evolução.
11	Eficiência e Baixa Latência	Parcialmente Satisfeito	O tempo de resposta é adequado, mas há espaço para otimização na velocidade de processamento.

6.2. Limitações

Apesar dos progressos alcançados no desenvolvimento do protótipo MANCAPT, algumas limitações foram identificadas e devem ser consideradas para futuras melhorias e otimizações do sistema. O desempenho do modelo melhora com um maior volume de dados, fazendo com que o processo de *data augmentation* seja necessário para garantir a eficiência.

A identificação do autor afeta diretamente a transcrição, pois o modelo carregado depende da autoria atribuída. Erros na identificação levam a um aumento nos erros de transcrição, especialmente em casos de caligrafias semelhantes. Apesar da elevada precisão, em algumas situações, manuscritos parecidos geraram confusão na atribuição do autor.

A qualidade das digitalizações também influencia a precisão do reconhecimento, sendo mais desafiador em documentos com baixo contraste, ruído ou escrita muito fina. Embora técnicas de pré-processamento de imagens ajudem a mitigar esses problemas, ainda há impacto na performance.

Os modelos personalizados superiorizaram-se aos modelos genéricos do *Azure AI Document Intelligence*, mas foram treinados para um conjunto específico de caligrafias. Quando aplicado a novas escritas, a precisão pode reduzir, exigindo ajustes. Além disso, o processo de treino exige recursos computacionais elevados, impactando o desempenho em sistemas com *hardware* limitado.

Durante os testes, verificou-se que variações na escrita de um mesmo autor, como inclinação ou pressão da caneta, podem afetar a precisão. Futuros desenvolvimentos devem integrar maior flexibilidade na adaptação a variações intra-autor. Além disso, a interoperabilidade com sistemas externos não foi exaustivamente testada.

As limitações identificadas não comprometem a eficácia do modelo, mas indicam oportunidades para melhoria, para garantir um sistema adaptável a diferentes cenários.

CAPÍTULO 7

Conclusões

7.1. Conclusão

O desenvolvimento do protótipo MANCAPT demonstrou que a personalização de modelos de OCR é um fator determinante para melhorar a precisão na identificação do autor e transcrição de manuscritos. Os resultados obtidos confirmam que a abordagem personalizada supera os modelos genéricos, proporcionando uma solução eficiente para cenários específicos. Ao longo deste trabalho foi possível comprovar que o protótipo MANCAPT atingiu os seus objetivos:

- **Objetivo 1:** Identificar o autor dos documentos manuscritos através da caligrafia
    - A identificação do autor através da caligrafia foi validada, onde atingiu 98% de precisão na identificação de autores;
- **Objetivo 2:** Transcrever textos manuais para texto digital
    - O sistema alcançou com sucesso a transcrição de textos manuscritos para formato digital, reduzindo a taxa de erro de caracteres (CER) para 6%, face aos 11,3% do modelo genérico *Azure AI Document Intelligence*;
    - O alinhamento de manuscritos e impressos na proporção 70% 30% foi essencial para otimizar o desempenho do modelo.

A criação deste protótipo, envolveu a necessidade de criação de um conjunto de dados, nomeado MANCAPT dataset, construído especificamente para treinar e validar o modelo de reconhecimento de autores e transcrição de manuscritos. Esse *dataset* inclui imagens manuscritas e impressas de todos os participantes na recolha de amostras e da demonstração e poderá ser disponibilizado para a comunidade científica.

Contudo, algumas limitações foram identificadas, nomeadamente a dependência de grandes volumes de dados para maximizar a precisão, a dificuldade em diferenciar autores com caligrafias muito semelhantes e a necessidade de hardware robusto para o treino eficiente do modelo.

No geral, os objetivos foram alcançados, demonstrando a viabilidade de um sistema OCR altamente personalizado e capaz de atender a necessidades específicas que soluções genéricas não conseguem suprir.

7.2. Trabalho Futuro

Embora os resultados obtidos sejam positivos e demonstrem que o MANCAPT pode acrescentar valor à investigação em sistemas OCR, ainda existem diversos aspetos a melhorar. A identificação do autor pode ser otimizada com técnicas mais avançadas de classificação, reduzindo erros em caligrafias semelhantes. Além disso, é essencial que o sistema seja capaz de atribuir uma classificação própria a caligrafias desconhecidas, evitando enviesamentos nos resultados. A implementação de técnicas de aprendizagem auto-supervisionada poderá diminuir a necessidade de anotação manual, tornando o modelo mais eficiente e autónomo.

Para possibilitar a expansão internacional do MANCAPT, será necessário adaptá-lo a múltiplos idiomas e estilos de escrita. A integração com sistemas externos, através do desenvolvimento de uma API permitiria que o sistema fosse incorporado em diferentes aplicações. Isto traria benefícios significativos, mas exigiria igualmente a criação de mecanismos robustos de segurança e privacidade para proteção dos dados.

A consolidação dos avanços apresentados neste trabalho abre caminho para sistemas OCR adaptáveis e orientados para a personalização, garantindo uma transição mais precisa da escrita manuscrita para o formato digital.

REFERÊNCIAS BIBLIOGRÁFICAS

[1] A. Chahi, Y. El merabet, Y. Ruichek, and R. Touahni, “WriterINet: a multi-path deep CNN for offline text-independent writer identification,” *International Journal on Document Analysis and Recognition*, vol. 26, no. 2, 2023, doi: 10.1007/s10032-022-00418-3.

[2] I. Kavinda and H. Fernando, “Handwritten Prescription Recognition Using VGG Based Architecture with Bi-LSTM,” in *2024 International Research Conference on Smart Computing and Systems Engineering (SCSE)*, 2024, pp. 1–6. doi: 10.1109/SCSE61872.2024.10550875.

[3] S. Kulkarni, R. Madurwar, R. Narlawar, A. Pandya, and N. Gawande, “Digitization of Physical Notes: A Comprehensive Approach Using OCR, CNN, RNN, and NMF,” in *2023 7th International Conference On Computing, Communication, Control And Automation (ICCUBEA)*, 2023, pp. 1–5. doi: 10.1109/ICCUBEA58933.2023.10391967.

[4] G. C. Lacerda and R. C. S. Vasconcelos, “Character Identification in Images Extracted from Portuguese Manuscript Historical Documents,” in *Proceedings of the International Joint Conference on Computer Vision, Imaging and Computer Graphics Theory and Applications*, 2024. doi: 10.5220/0012353900003660.

[5] Z. Zhelev and S. Kostova, “Investigating the Application of Digital Tools for Information Management in Financial Control: Evidence from Bulgaria,” *Journal of Risk and Financial Management*, vol. 17, no. 4, 2024, doi: 10.3390/jrfm17040165.

[6] B. J. Ngereja, “Exploring the Role of Soft Factors in Enhancing Digitalization Project Success A Multi-layer Perspective Emphasizing the Influence of Individual, Team, and Organizational Learning,” 2024. Accessed: Apr. 25, 2024. [Online]. Available: https://ntnuopen.ntnu.no/ntnu-xmlui/handle/11250/3126000

[7] C. Makó and J. Pap, “Algorithmic management and democracy at work in Hungary Policy brief,” 2024. Accessed: Apr. 25, 2024. [Online]. Available: https://ddd.uab.cat/pub/infpro/2024/290694/INCODING_Policy_Brief_Hungary_def.pdf

[8] T. Limpo, R. A. Alves, and V. Connelly, “Examining the transcription-writing link: Effects of handwriting fluency and spelling accuracy on writing performance via planning and translating in middle grades,” *Learn Individ Differ*, vol. 53, pp. 26–36, Jan. 2017, doi: 10.1016/J.LINDIF.2016.11.004.

[9] S. Araújo, M. Domingues, and T. Fernandes, “From Hand to Eye: a Meta-Analysis of the Benefit from Handwriting Training in Visual Graph Recognition,” *Educ Psychol Rev*, vol. 34, no. 3, 2022, doi: 10.1007/s10648-021-09651-4.

[10] F. Heinrichs, “Unstructured interaction: Integrating informal handwritten knowledge into business processes,” in *LWA 2009 - Workshop-Woche: Lernen-Wissen-Adaptivitat - Learning, Knowledge, and Adaptivity*, 2009.

[11] C. Adak, B. B. Chaudhuri, and M. Blumenstein, “An empirical study on writer identification and verification from intra-variable individual handwriting,” *IEEE Access*, vol. 7, 2019, doi: 10.1109/ACCESS.2019.2899908.

[12] E. Breci, L. Guarnera, and S. Battiato, “Innovative Methods for Non-Destructive Inspection of Handwritten Documents,” in *ICASSP 2024 - 2024 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*, 2024, pp. 4825–4829. doi: 10.1109/ICASSP48485.2024.10448383.

[13] A. Benfell, “Modeling functional requirements using tacit knowledge: a design science research methodology informed approach,” *Requir Eng*, vol. 26, no. 1, 2021, doi: 10.1007/s00766-020-00330-4.

[14] S. S. Devi and T. Amitha, “Offline handwritten writer independent Tamil character recognition,” in *2014 International Conference on Information Communication and Embedded Systems, ICICES 2014*, 2015. doi: 10.1109/ICICES.2014.7033838.

[15] S. Katoch, M. Rakhra, and D. Singh, “Recognition Of Handwritten English Character Using Convolutional Neural Network,” in *2022 4th International Conference on Artificial Intelligence and Speech Technology (AIST)*, 2022, pp. 1–6. doi: 10.1109/AIST55798.2022.10064860.

[16] L. Navya, MD. F. Ali, K. P. Sai, K. Shyam, and A. Ramesh, “Handwritten Text Recognition Using Deep Learning Techniques,” in *2023 Annual International Conference on Emerging Research Areas: International Conference on Intelligent Systems (AICERA/ICIS)*, 2023, pp. 1–5. doi: 10.1109/AICERA/ICIS59538.2023.10420040.

[17] G. G, M. Apsar, Neha, J. Shulamite, and N. S. Maddipudi, “Automation of Sign and Photo Detection Using AI and OCR,” in *2023 4th IEEE Global Conference for Advancement in Technology (GCAT)*, 2023, pp. 1–6. doi: 10.1109/GCAT59970.2023.10353460.

[18] G. G. Rajput and S. B. Ummapure, “Script identification from handwritten documents using SIFT method,” in *2017 IEEE International Conference on Power, Control, Signals and Instrumentation Engineering (ICPCSI)*, 2017, pp. 520–526. doi: 10.1109/ICPCSI.2017.8392348.

[19] C. Adak, B. B. Chaudhuri, C.-T. Lin, and M. Blumenstein, “Intra-Variable Handwriting Inspection Reinforced With Idiosyncrasy Analysis,” *IEEE Transactions on Information Forensics and Security*, vol. 15, pp. 3567–3579, 2020, doi: 10.1109/TIFS.2020.2991833.

[20] B. M. Garlapati and S. R. Chalamala, “A System for Handwritten and Printed Text Classification,” in *Proceedings - 2017 UKSim-AMSS 19th International Conference on Modelling and Simulation, UKSim 2017*, 2018. doi: 10.1109/UKSim.2017.37.

[21] M. Chethan, R. Anirudh, M. K. Rani, and S. R. Dey, “A Novel Segmentation-Free Approach for Handwritten Sentence Recognition,” in *Lecture Notes in Electrical Engineering*, 2023. doi: 10.1007/978-981-19-7346-8_55.

[22] J. and R. K. B. and R. Y. P. and P. K. G. S. P. Raja Subramanium R. and Visal, “IAM Dataset-Based Author Identification via Convolutional Neural Networks,” in *Smart Trends in Computing and Communications*, C. and J. A. Senjyu Tomonobu and So–In, Ed., Singapore: Springer Nature Singapore, 2024, pp. 81–93.

[23] S. D. Kale and R. S. Prasad, “A Systematic Review on Author Identification Methods,” *International Journal of Rough Sets and Data Analysis*, vol. 4, no. 2, 2017, doi: 10.4018/ijrsda.2017040106.

[24] M. J. Page *et al.*, “The PRISMA 2020 statement: An updated guideline for reporting systematic reviews,” 2021. doi: 10.1136/bmj.n71.

[25] J. H. Kim, “Improvement of Inceptionv3 model classification performance using chest x-ray images,” *J Mech Med Biol*, vol. 22, no. 8, Oct. 2022, doi: 10.1142/S0219519422400322.

[26] L. Ali, F. Alnajjar, H. Al Jassmi, M. Gochoo, W. Khan, and M. A. Serhani, “Performance evaluation of deep CNN-based crack detection and localization techniques for concrete structures,” *Sensors*, vol. 21, no. 5, 2021, doi: 10.3390/s21051688.

[27] Microsoft, “Build and Deploy Custom Models,” Microsoft Learn. Accessed: Oct. 11, 2024. [Online]. Available: https://learn.microsoft.com/pt-br/azure/architecture/ai-ml/architecture/build-deploy-custom-models

[28] C. Khosla and B. S. Saini, “Enhancing Performance of Deep Learning Models with different Data Augmentation Techniques: A Survey,” in *Proceedings of International Conference on Intelligent Engineering and Management, ICIEM 2020*, 2020. doi: 10.1109/ICIEM48762.2020.9160048.

[29] F. Zhuang *et al.*, “A Comprehensive Survey on Transfer Learning,” 2021. doi: 10.1109/JPROC.2020.3004555.

[30] A. Krizhevsky, I. Sutskever, and G. E. Hinton, “ImageNet classification with deep convolutional neural networks,” *Commun ACM*, vol. 60, no. 6, pp. 84–90, Jun. 2017, doi: 10.1145/3065386.

[31] F. M. Talaat, M. Shams, S. Adel Gamel, and H. Zaineldin, “DeepLeaf: an optimized deep learning approach for automated recognition of grapevine leaf diseases,” *Neural Comput Appl*, pp. 1–25, Mar. 2025, doi: 10.1007/s00521-025-11038-3.

[32] K. Kettunen, M. Koistinen, and J. Kervinen, “Ground truth ocr sample data of finnish historical newspapers and journals in data improvement validation of a re-ocring process,” *LIBER Quarterly*, vol. 30, no. 1, 2020, doi: 10.18352/lq.10322.

[33] R. Likert, S. Roslow, and G. Murphy, “A simple and reliable method of scoring the Thurstone attitude scales,” *Pers Psychol*, vol. 46, no. 3, 1993, doi: 10.1111/j.1744-6570.1993.tb00893.x.