



Universidade Estadual de Maringá - Departamento de Informática  
Bacharelado em Informática



Aluno: Yuri Pires Alves

# **Desenvolvimento de uma Ferramenta Baseada em Características Grafométricas para Identificação de Manuscritos**

Orientadora: Prof<sup>a</sup>. Dr<sup>a</sup>. Aline Maria Malachini Miotto Amaral

Trabalho de Conclusão de Curso

2025

# Conteúdo

**01 - Introdução**

**02 - Fundamentação teórica**

**03 - Materiais e métodos**

**04 - Desenvolvimento da ferramenta**

**05 - Experimentos**

**06 - Resultados**

**07 - Conclusão**

# 01 - Introdução

# Contextualização

A identificação de autoria em documentos manuscritos é uma tarefa crucial em áreas como a criminalística e a ciência forense. Ela é essencial para:

elucidação de crimes

resolução de disputas legais

validação de assinaturas

entre outros

A **análise grafométrica**, tradicionalmente realizada por peritos em documentoscopia, consiste na análise minuciosa de características da escrita, como:

forma das letras

inclinação

pressão

espaçamento

entre outros

Ela tem o objetivo de determinar a autoria de um documento questionado.

# Problemas

grande variabilidade da escrita humana

similaridade entre diferentes autores

subjetividade intrínseca ao processo de análise humana, especialmente quando lida com um grande volume de documentos

Todos esses fatores **aumentam a complexidade** da tarefa e podem impactar o **tempo de resolução de casos**, bem como a **precisão dos resultados**.

# Solução proposta

Este trabalho propõe o desenvolvimento de uma ferramenta para automatizar a identificação de manuscritos com o uso de características grafométricas.

## Objetivos

Refatorar algoritmos anteriores que implementam a extração de características grafométricas, com foco na **inclinação axial**, de forma modular, possibilitando assim a adição de novas características em estudos futuros.

Elaborar a interface web para utilização da ferramenta.

Realizar experimentos para avaliação da ferramenta desenvolvida.

Analizar os resultados obtidos com os experimentos realizados.

# **02 - Fundamentação teórica**

# Escrita como elemento biométrico

A **documentoscopia**, ramo da criminalística dedicado à análise de documentos, utiliza características específicas da escrita para verificar a autenticidade e determinar a autoria de manuscritos.

Cada indivíduo desenvolve um estilo único de escrita, influenciado por fatores:

fisiológicos

psicológicos

sociais

culturais

Esses fatores geram nuances e padrões exclusivos, tornando a escrita uma verdadeira “impressão digital gráfica”.

Essa singularidade da escrita humana a coloca como um elemento biométrico com alto potencial para a identificação de indivíduos.

# Princípios fisiológicos da escrita

Saudek (1978) descreve alguns princípios fisiológicos que regem a escrita e que contribuem para a individualidade do gesto gráfico:

Habitualidade do gesto

Ritmo muscular e relaxamento

Impacto da fadiga na fluidez e uniformidade

Modificações na escrita em situações de estresse ou cansaço

A combinação desses princípios, atuando de forma singular em cada indivíduo, contribui para a formação de um estilo de escrita único, que pode ser utilizado como base para a identificação de autoria.

# Características grafométricas

Elementos mensuráveis da escrita que ajudam a identificar os padrões únicos presentes no traçado manual de cada indivíduo.

São utilizadas na análise forense para verificar a autenticidade de documentos e determinar a autoria de manuscritos.

## Globais

Representam aspectos gerais da escrita, extraídos de um documento, parágrafo ou linha.

### Inclinação axial

Ângulo de inclinação da escrita.

### Espaçamento entre linhas

Distância média entre as linhas de texto.

### Tamanho das margens

Dimensão das margens superior, inferior, esquerda e direita.

### Alinhamento do texto

Padrão de alinhamento em relação à linha de base.

## Locais

Referem-se a características específicas de palavras ou letras individuais.

### Forma das letras

Presença de laços, ornamentos ou distorções.

### Pressão da escrita

Intensidade do traçado, que reflete a força aplicada.

### Ligações entre letras

Forma e fluidez na conexão entre letras.

### Detalhes gráficos

Pontos, acentos e outros elementos menores que variam entre indivíduos.

# Características grafométricas

Elementos mensuráveis da escrita que ajudam a identificar os padrões únicos presentes no traçado manual de cada indivíduo.

São utilizadas na análise forense para verificar a autenticidade de documentos e determinar a autoria de manuscritos.

## Globais

Representam aspectos gerais da escrita, extraídos de um documento, parágrafo ou linha.

### Inclinação axial

Ângulo de inclinação da escrita.

### Espaçamento entre linhas

Distância média entre as linhas de texto.

### Tamanho das margens

Dimensão das margens superior, inferior, esquerda e direita.

### Alinhamento do texto

Padrão de alinhamento em relação à linha de base.

## Locais

Referem-se a características específicas de palavras ou letras individuais.

### Forma das letras

Presença de laços, ornamentos ou distorções.

### Pressão da escrita

Intensidade do traçado, que reflete a força aplicada.

### Ligações entre letras

Forma e fluidez na conexão entre letras.

### Detalhes gráficos

Pontos, acentos e outros elementos menores que variam entre indivíduos.

# Características grafométricas

Elementos mensuráveis da escrita que ajudam a identificar os padrões únicos presentes no traçado manual de cada indivíduo.

São utilizadas na análise forense para verificar a autenticidade de documentos e determinar a autoria de manuscritos.

## Globais

Representam aspectos gerais da escrita, extraídos de um documento, parágrafo ou linha.

### Inclinação axial

Ângulo de inclinação da escrita.

### Espaçamento entre linhas

Distância média entre as linhas de texto.

### Tamanho das margens

Dimensão das margens superior, inferior, esquerda e direita.

### Alinhamento do texto

Padrão de alinhamento em relação à linha de base.

## Locais

Referem-se a características específicas de palavras ou letras individuais.

### Forma das letras

Presença de laços, ornamentos ou distorções.

### Pressão da escrita

Intensidade do traçado, que reflete a força aplicada.

### Ligações entre letras

Forma e fluidez na conexão entre letras.

### Detalhes gráficos

Pontos, acentos e outros elementos menores que variam entre indivíduos.

# Características grafométricas

A importância das características grafométricas na identificação de autoria se justifica por sua sólida fundamentação científica, baseada em princípios e leis da escrita estabelecidos pela **grafoscopia**, uma área de estudo consolidada e reconhecida pela comunidade científica.

Além disso, a interpretabilidade dessas características, por serem intuitivas e facilmente compreendidas por peritos e juízes, facilita a argumentação em processos judiciais.

As características grafométricas também demonstram robustez, sendo muitas vezes resistentes a tentativas de disfarce, pois refletem hábitos profundamente enraizados na escrita do indivíduo.

Sua aplicabilidade a documentos existentes, permitindo a análise de manuscritos já produzidos, torna essa abordagem útil em investigações criminais e processos judiciais.

## Características grafométricas

No entanto, a extração e a análise automatizada de características grafométricas apresentam desafios, como a complexidade algorítmica, que exige o desenvolvimento de algoritmos robustos para extrair essas características de imagens de manuscritos, demandando conhecimento em processamento de imagens, visão computacional e inteligência artificial.

Outro desafio reside na variação intrapessoal, pois a escrita de um mesmo indivíduo pode variar dependendo de fatores como o estado emocional, a fadiga, a velocidade de escrita e a intenção de disfarce.

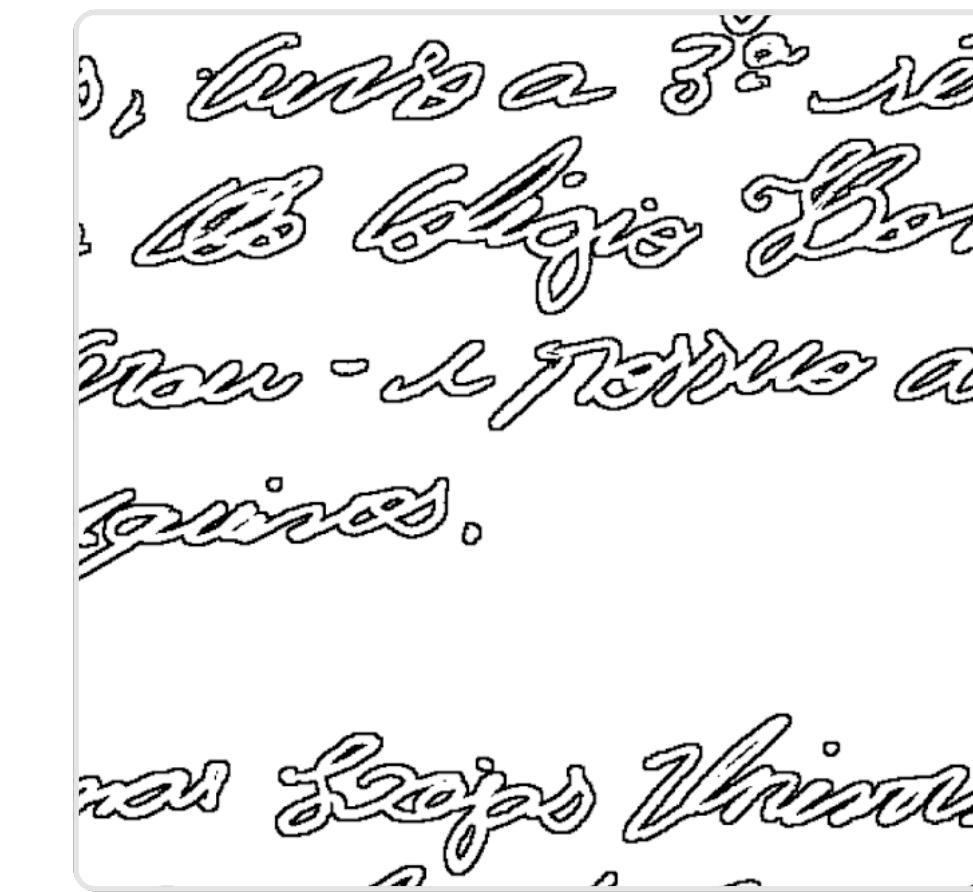
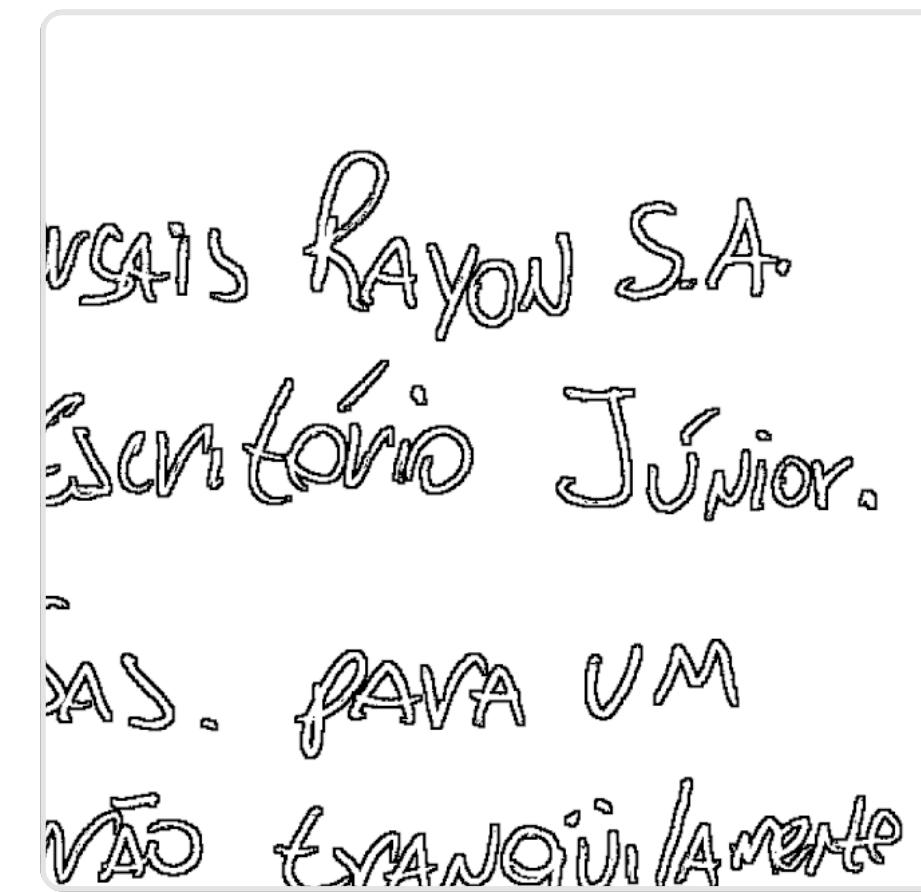
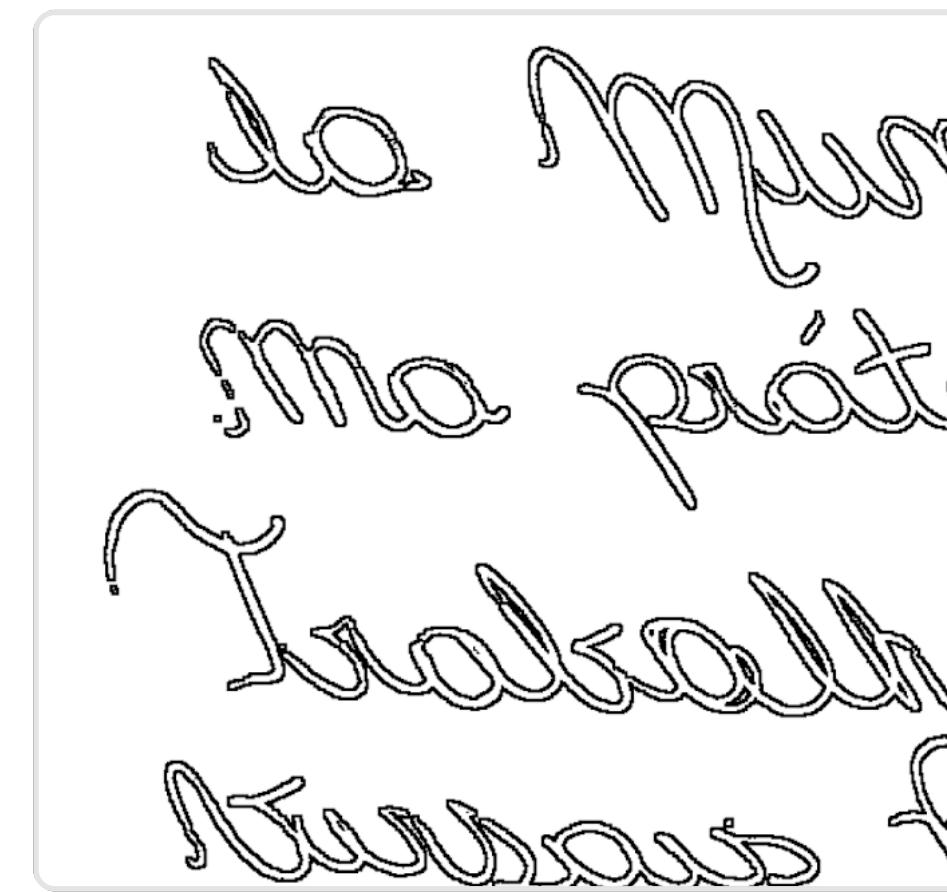
# Inclinação axial

Ângulo de inclinação da escrita em relação ao eixo vertical, sendo o eixo horizontal representado por uma linha de base imaginária.

## Alta capacidade discriminatória

Amplamente utilizada por peritos em suas análises

## Base sólida em estudos prévios



Devido à sua relevância prática e comprovada eficácia na identificação de autoria, a inclinação axial foi selecionada como a característica a ser implementada neste trabalho.

# **03 - Materiais e métodos**

# Recursos computacionais

## Hardware

### Processador

Intel Core i5 12450H

### Placa de vídeo

Nvidia RTX 3050 4GB VRAM

### Memória RAM

16GB

### Sistema operacional

Ubuntu 22.04 (WSL)

## Software

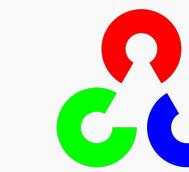
### Backend



Python



Scikit-learn

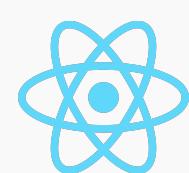


OpenCV

### Frontend



TypeScript



React



Next.js

# Base de dados

A base de dados utilizada para a realização dos experimentos de avaliação da ferramenta desenvolvida foi a **Base de Cartas Forenses Brasileira**, também conhecida como **Cartas Forenses PUCPR**.

Desenvolvida considerando as particularidades da língua portuguesa, como a presença de acentos e do cedilha

Contém 1800 exemplares de cartas

600 autores distintos, cada autor contribuiu com 03 cartas

Cartas redigidas em folha A4 não pautada

Digitalizadas em 300 dpi

Disponíveis em escala de cinza (256 níveis de cinza)

De  
FERNANDO QUINTAS ZANON  
RUA LUIZ KIRK WA KEREZ, 87 - AP. 300  
XENAPOLIS, NOVA YORKA 14506-159

PARA  
Dr. Onório Bob Grant

Soube, através de publicação pela imprensa local, que V. SAs.  
necessitam de um funcionário na Seção de Correspondência do  
Departamento fiscal. Venho, portanto, candidatar-me a esta vaga.  
Sou brasileiro, solteiro, com 18 anos, curso a 3<sup>a</sup> série do Curso  
Técnico de Contabilidade do Colégio Horácio Alves - Escola  
Municipal de 2º Grau - e posso alguma prática de datilografia  
e Arquivos.

Trabalhei durante dois anos nas Lojas Universais RAYON S.A.  
onde exercei as funções de Auxiliar de Escritório Júnior.  
Inicialmente, coloco-me à disposição de V. SAs. para um  
período de experiência, quando, então, poderão tranquilamente  
avaliar minhas aptidões.

Na expectativa de uma resposta apresento-lhes cordiais  
SAUDAÇÕES

Fernando Zanon

# Metodologia

Revisão sistemática da literatura

Refatoração de algoritmos

Preparação dos dados

Extração da característica grafométrica

Desenvolvimento da interface

# **04 - Desenvolvimento da ferramenta**

# Preparação da base de dados

Para garantir a generalização do modelo de identificação de autoria e avaliar seu desempenho em dados não utilizados durante o treinamento, a base de dados foi dividida em dois conjuntos.

Conjunto de Treinamento

**1200**

exemplares

Conjunto de Teste

**600**

exemplares

Essa divisão visa simular um cenário real, onde o modelo é treinado com dados de autores conhecidos e posteriormente avaliado em documentos de autoria desconhecida.

# Módulos da ferramenta

## Backend

Responsável por todas as etapas de processamento e análise dos dados, incluindo:

Pré-processamento das imagens

Extração da característica grafométrica

Treinamento e teste dos modelos de identificação de autoria (SVM e Random Forest)

Envio dos resultados para o frontend

## Frontend

Responsável pela interface com o usuário, com as seguintes funcionalidades:

Configuração dos parâmetros para o treinamento dos modelos SVM e Random Forest

Visualização dos resultados da análise de autoria

# Criação da API

A comunicação entre o frontend e o backend da ferramenta é realizada por meio de uma API REST, desenvolvida em Python utilizando a biblioteca FastAPI.

A estrutura da API é composta por uma rota principal, responsável por receber os parâmetros de configuração do modelo de identificação de autoria. Após receber a requisição com os parâmetros, a API aciona os módulos responsáveis pelas seguintes etapas:

Processamento das imagens da base de dados

Extração da inclinação axial

Criação e classificação dos modelos

Em seguida, a API retorna os resultados do treinamento para o frontend. Esses resultados incluem métricas de desempenho do modelo, como a acurácia, permitindo ao usuário avaliar a qualidade do modelo gerado.

# Vantagens da API

Modularidade

Flexibilidade

Padronização

Eficiência

A API, portanto, é um componente crucial para o funcionamento da ferramenta, viabilizando a comunicação eficiente e a troca de informações entre o frontend e o backend, garantindo a usabilidade e a robustez da ferramenta como um todo.

# Processamento de imagens

**Binarização:** processo que simplifica a representação visual, convertendo uma imagem em escala de cinza de 256 níveis para uma imagem binária, composta apenas por pixels pretos e brancos.

De

Fernando Quintas Zanon

Rua Luiz Kirt Warez, 87 - Ap. 300  
Xentapolis, Nova York 14506-158

PAVA

Dr. Onório Bob Grant

Soube, através de publicação pela imprensa local, que V. Srs.

Necessitam de um funcionário na Seção de Correspondência do  
Departamento Federal. Venho, portanto, candidatar-me a esta vaga.

Sou brasileiro, solteiro, com 18 anos, curso a 3<sup>a</sup> série do Curso

# Processamento de imagens

Dilatação: modifica a imagem de um manuscrito deixando o traçado do autor mais espesso.

De  
Fernando Quintas Zanon  
RUA Luiz Kirt wa Kerez, 87 - Ap. 300  
Xentropolis, Nova York 14506-158

Para  
Dr. Onório Bob Grant

Soube, através de publicação pela imprensa local, que V. Sua  
Necessitam de um funcionário na Seção de Correspondência do  
Departamento fiscal. Venho, portanto, candidatar-me a esta vaga.  
Sou brasileiro, solteiro, com 18 anos, curso a 3<sup>a</sup> série do Curso

# Processamento de imagens

**Erosão:** modifica a imagem de um manuscrito deixando o traçado do autor mais fino.



# Processamento de imagens

**Borda:** Após a geração da imagem dilatada e imagem erodida, estas são sobrepostas e é feita a subtração dos pixels compatíveis nas duas imagens resultando na imagem de borda.

De  
Fernando QUINTAS ZANON  
RUA LUIZ KIRK DA CEREZ, 87 - Ap. 300  
Xenopolis, Nova York 14506-158

PARA  
Dr. Onório Bob Grant

Soube, através de publicação pela imprensa local, que V. SAs.  
Necessitam de um funcionário na Seção de Correspondência do  
Departamento fiscal. Venho, portanto, considerar-me à essa VAGA  
Sou brasileiro, solteiro, com 18 anos, curso à 3<sup>a</sup> série do Curso

# Extração da característica Inclinação Axial

Para este trabalho, a abordagem escolhida para extração da característica foi a técnica da **distribuição de borda direcional**.

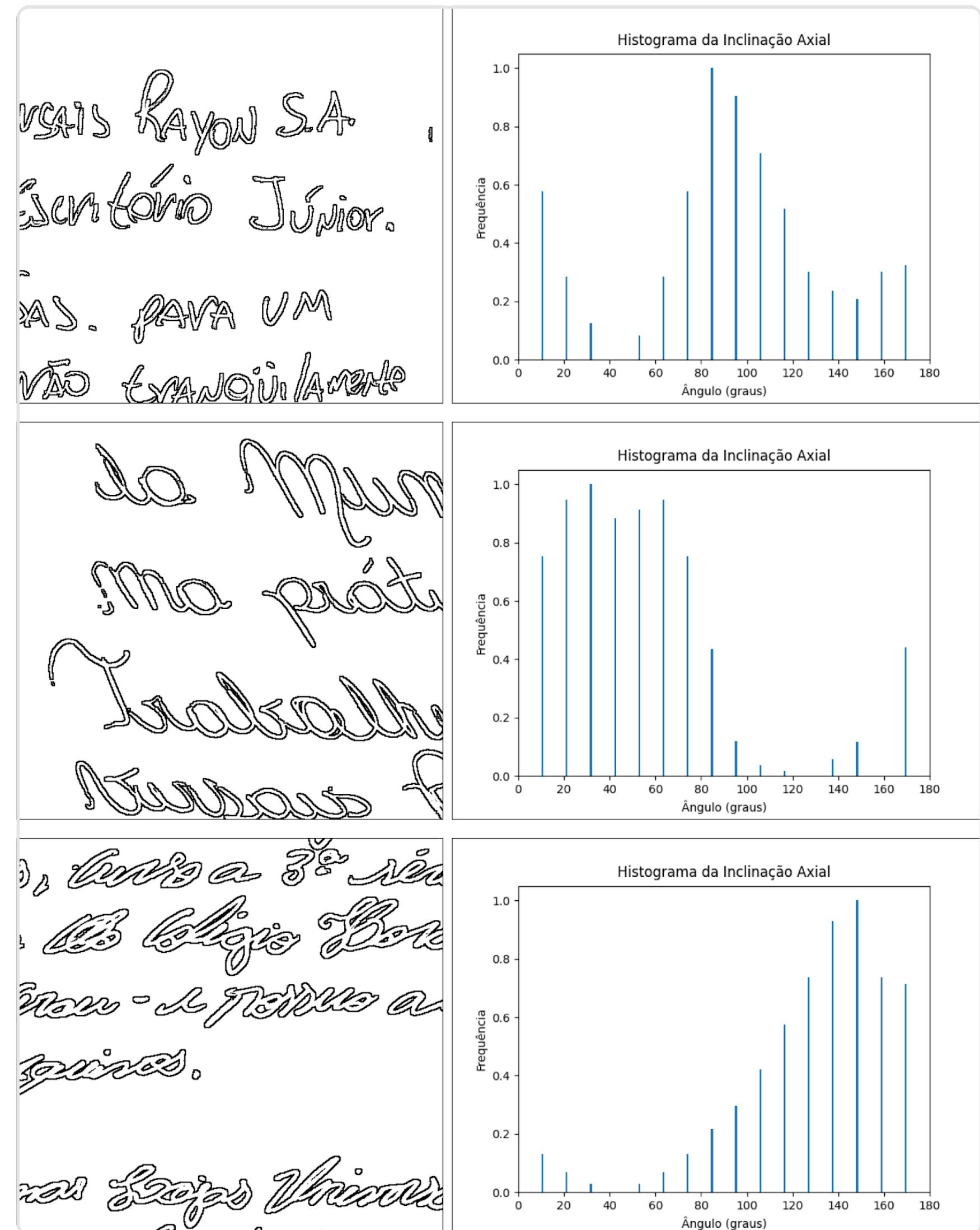
Segundo Baranowski (2005), nesta técnica, consideram-se as bordas, pois obtiveram melhores resultados na discriminação da inclinação axial do autor em relação à inclinação extraída diretamente do traçado e por serem mais finas, reduzem a influência da espessura do traçado sobre o cálculo.

A imagem pré-processada é então percorrida considerando o pixel da borda do traçado no centro de um elemento estruturante retangular. Em seguida, são verificados os fragmentos de borda em todas as direções, partindo deste pixel central e conferindo os pixels posteriores com um operador lógico AND, finalizando nas extremidades do elemento estruturante apenas se houver a presença de um fragmento de borda inteiro.

Ou seja, se todos os pixels vizinhos forem pretos, considera-se o fragmento de borda e computa-se a posição do fragmento em um vetor de posições.

Este vetor é normalizado para garantir que todas as características tenham a mesma escala em um intervalo de  $[0, 1]$ .

Após a normalização, o vetor resultante representa a distribuição da inclinação axial do manuscrito, com valores próximos a 1 indicando uma maior frequência de pixels naquela direção (e, portanto, uma maior probabilidade de inclinação) e valores próximos a 0 indicando uma menor frequência.



# Preparação dos dados para os modelos de classificação

Após a extração da característica inclinação axial e a normalização dos vetores de distribuição de borda direcional, são gerados dois arquivos CSV:

Início do arquivo CSV de treino

autor	inclinacao_0	inclinacao_1	inclinacao_2	inclinacao_3
a001	0.3658632150674197	0.5968195206312319	0.28402734617775016	0.12490880103764153
a001	0.32954545454545453	0.6061847988077497	0.28808896022010777	0.12669093201880088
a002	1.0	0.23980191265459783	0.09890470836438976	0.03759616963755574
a002	1.0	0.25756237032036455	0.11871457077902094	0.04636202086787321
a003	0.4480336903140347	0.5429564000349899	0.2454669282581258	0.1015583019881784
a003	0.4713869765163678	0.5331981822634629	0.23192467822253668	0.09666929712373126
a004	0.47907760324864396	0.4129069325828472	0.18490782849185405	0.07255889140998549
a004	0.4782795698924731	0.4435330261136713	0.19481822836661547	0.07524833589349718

Início do arquivo CSV de teste

autor	inclinacao_0	inclinacao_1	inclinacao_2	inclinacao_3
a001	0.34759691057505765	0.5894188415876618	0.2771697353490245	0.1226740851910148
a002	1.0	0.26092990094976193	0.12023527614391567	0.048251979731622235
a003	0.43458008811017557	0.5346444666052336	0.23537282722579178	0.0970600487256845
a004	0.5229634854680318	0.4491959749148582	0.20063704797698692	0.07717318226405528

No caso, como o vetor de distribuição possui 17 posições, cada linha do CSV terá 17 colunas representando esses valores. Uma coluna a mais é adicionada para identificar a classe (autor) à qual o manuscrito pertence.

# Criação e treinamento dos modelos

O objetivo principal é capacitar a ferramenta a identificar padrões nos dados, permitindo a identificação automática da autoria. Para alcançar este objetivo, a ferramenta oferece a flexibilidade de utilizar dois algoritmos, o **SVM** e o **Random Forest**.

Além disso, é possível escolher a quantidade de autores que será utilizada, sendo esses autores escolhidos de forma aleatória.

Adicionalmente, em trabalhos futuros, a ferramenta poderá ser expandida para incorporar outros modelos de aprendizado de máquina, enriquecendo ainda mais as opções de análise e adaptando-se às demandas específicas de cada pesquisa.

# SVM (Support Vector Machine)

No contexto da identificação de autoria em manuscritos, o SVM atua classificando manuscritos em diferentes classes, onde cada classe representa um autor.

A ideia central do SVM é encontrar o hiperplano ótimo que melhor separa as classes em um espaço multidimensional. Cada dimensão deste espaço corresponde a um valor da característica extraída, e cada manuscrito é representado como um ponto neste espaço.

O hiperplano ótimo é aquele que maximiza a margem, ou seja, a distância entre o hiperplano e os pontos de dados mais próximos de cada classe.

# SVM (Support Vector Machine)

Para utilizar o SVM, foi utilizada a implementação da biblioteca Scikit-learn.

O treinamento é realizado através do método **fit**, utilizando como entrada a matriz da característica de treinamento (`X_train`) e o vetor de rótulos correspondentes (`y_train`). A matriz `X_Train` é estruturada de forma que cada linha representa um manuscrito e cada coluna um valor da característica. Já o vetor `y_train` indica o autor associado a cada manuscrito em `X_train`.

# SVM com Grid Search

Visando não apenas construir um modelo funcional, mas também alcançar o desempenho ótimo, a técnica de Grid Search foi aplicada ao SVM.

Este método consiste em uma busca exaustiva dentro de um espaço predefinido de hiperparâmetros, testando diferentes combinações em busca daquela que maximiza a performance do modelo.

## kernel

Define a função matemática responsável por projetar os dados em um espaço de características de maior dimensionalidade, com o objetivo de facilitar a separação das classes. Os kernels avaliados foram: **linear**, **poly** e **rbf**.

## gamma

Este hiperparâmetro controla a influência de um único exemplo de treinamento na decisão do modelo. Valores baixos de gamma resultam em áreas de decisão mais amplas e suaves, enquanto valores altos criam áreas de decisão mais complexas e focadas em torno dos exemplos de treinamento.

## C

Parâmetro de regularização que atua como um contrapeso entre maximizar a margem entre as classes e minimizar o erro de classificação. Valores altos de C dão mais ênfase à classificação correta dos exemplos de treinamento, enquanto valores baixos buscam uma melhor capacidade de generalização.

## degree

Grau do polinômio para o kernel **poly**.

# Random Forest

Este algoritmo constrói uma floresta composta por várias árvores de decisão, onde cada árvore é treinada com uma amostra aleatória dos dados de treinamento.

Essa estratégia de aprendizado, é eficaz para evitar o problema de overfitting, quando o modelo se ajusta excessivamente aos dados de treinamento, comprometendo sua capacidade de generalização.

A implementação do Random Forest também utilizou-se da biblioteca Scikit-learn e o número de árvores foi definido como 100.

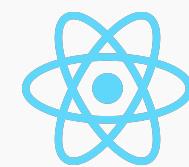
O treinamento do modelo Random Forest segue o mesmo padrão do SVM, utilizando o método **fit** com a matriz da característica de treinamento (`X_train`) e o vetor de rótulos (`y_train`) como entrada.

# Criação da interface

Para a visualização e utilização da ferramenta, foi desenvolvida uma interface web utilizando:



TypeScript



React



Next.js

Essa escolha se justifica pela versatilidade e amplo suporte da comunidade que essas tecnologias oferecem, permitindo a criação de uma interface moderna, responsiva e de fácil manutenção.

A interface tem como principal função enviar as configurações para o backend, que incluem o número de autores desejado e os modelos que serão utilizados para análise (SVM, SVM com Grid Search e/ou Random Forest). Após o processamento pelo backend, os resultados, como acurácia de cada modelo e os melhores parâmetros utilizados no Grid Search do SVM (caso o usuário tenha selecionado este modelo), são exibidos na interface.

## Manuscritus

Ferramenta baseada em características grafométricas para a identificação de manuscritos.

Ajuste o número de autores e selecione os modelos específicos para análise. Após definir as configurações, clique em **Obter resultados** para visualizar a acurácia de cada modelo.

### Número de Autores

Escolha um valor entre 20 e 200.

20



### Modelos

- SVM
- Random Forest

**Obter resultados**

## Manuscritus

Ferramenta baseada em características grafométricas para a identificação de manuscritos.

Ajuste o número de autores e selecione os modelos específicos para análise. Após definir as configurações, clique em **Obter resultados** para visualizar a acurácia de cada modelo.

### Número de Autores

Escolha um valor entre 20 e 200.

60



### Modelos

SVM

Random Forest

**Obter resultados**

### Acurácia:

SVM

**78.33%**

SVM (Grid Search)

**81.67%**

Random Forest

**78.33%**

### Melhores parâmetros (SVM):

C

**100**

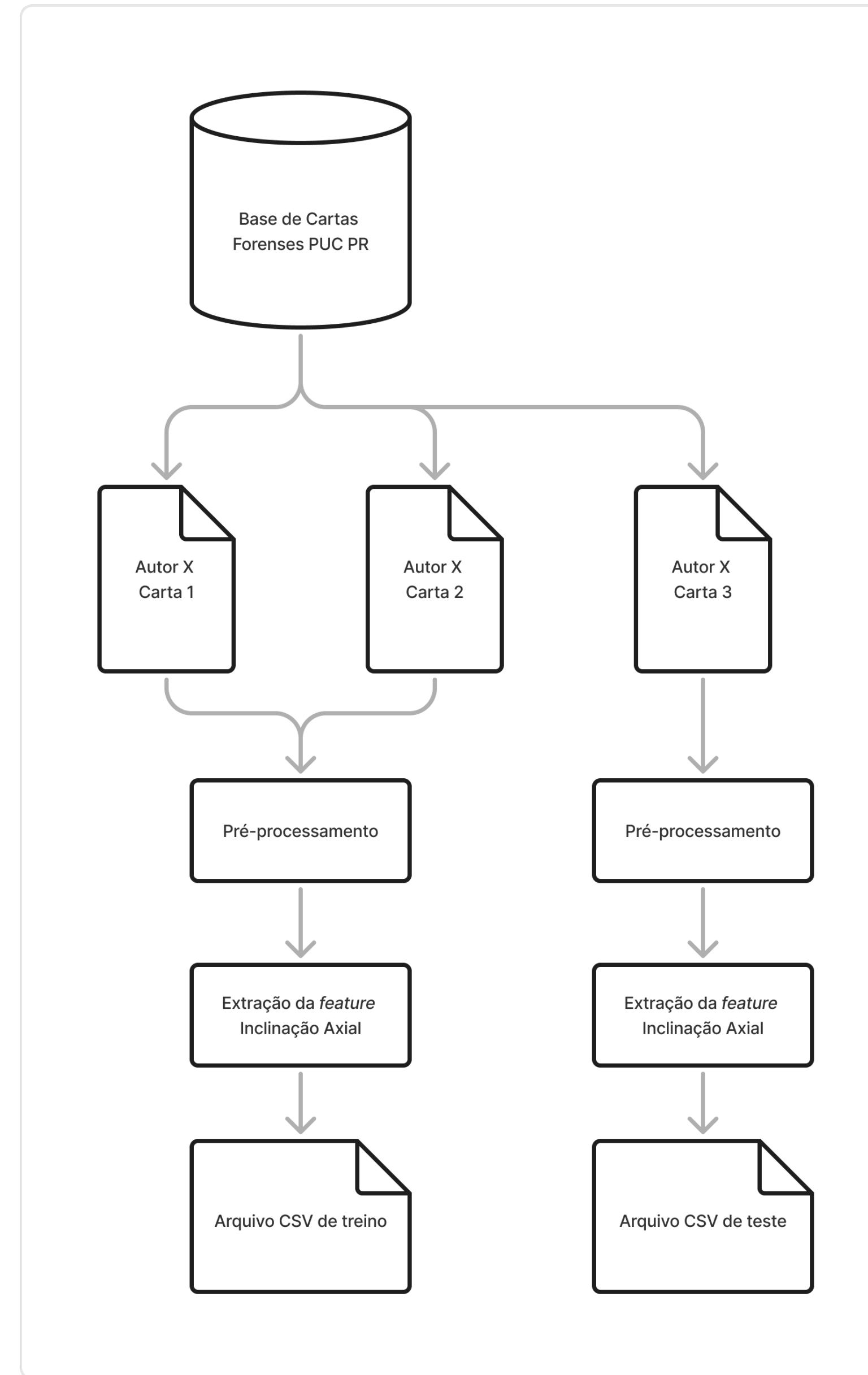
kernel

**linear**

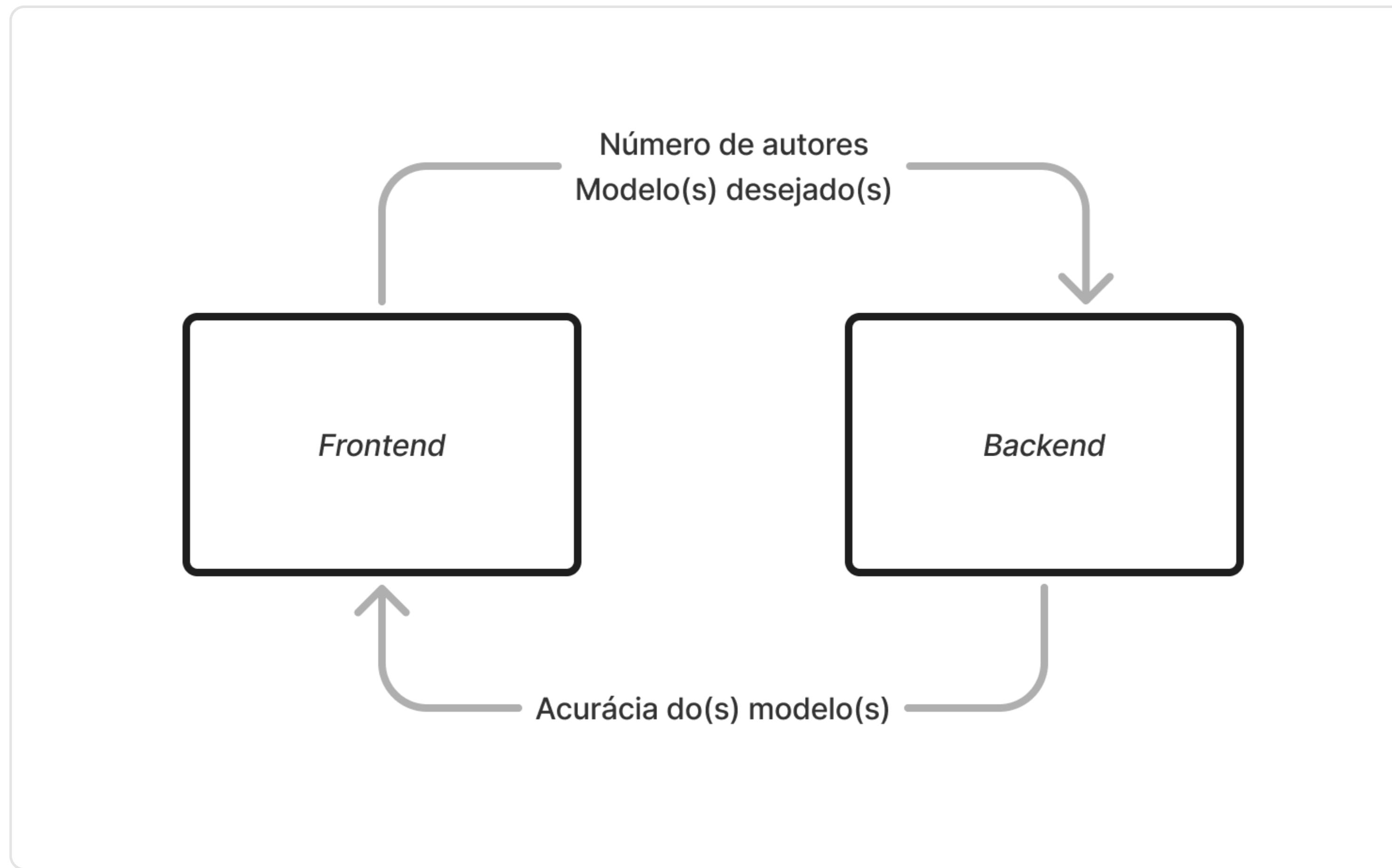
# **05 - Experimentos**

# Protocolo dos experimentos

Inicialmente, para cada autor presente na base de dados PUCPR, duas cartas foram alocadas ao conjunto de treinamento e uma carta ao conjunto de teste. O pré-processamento e a extração da inclinação axial foram realizados em cada imagem, e os vetores da característica resultantes foram salvos nos arquivos CSV de treino e teste.

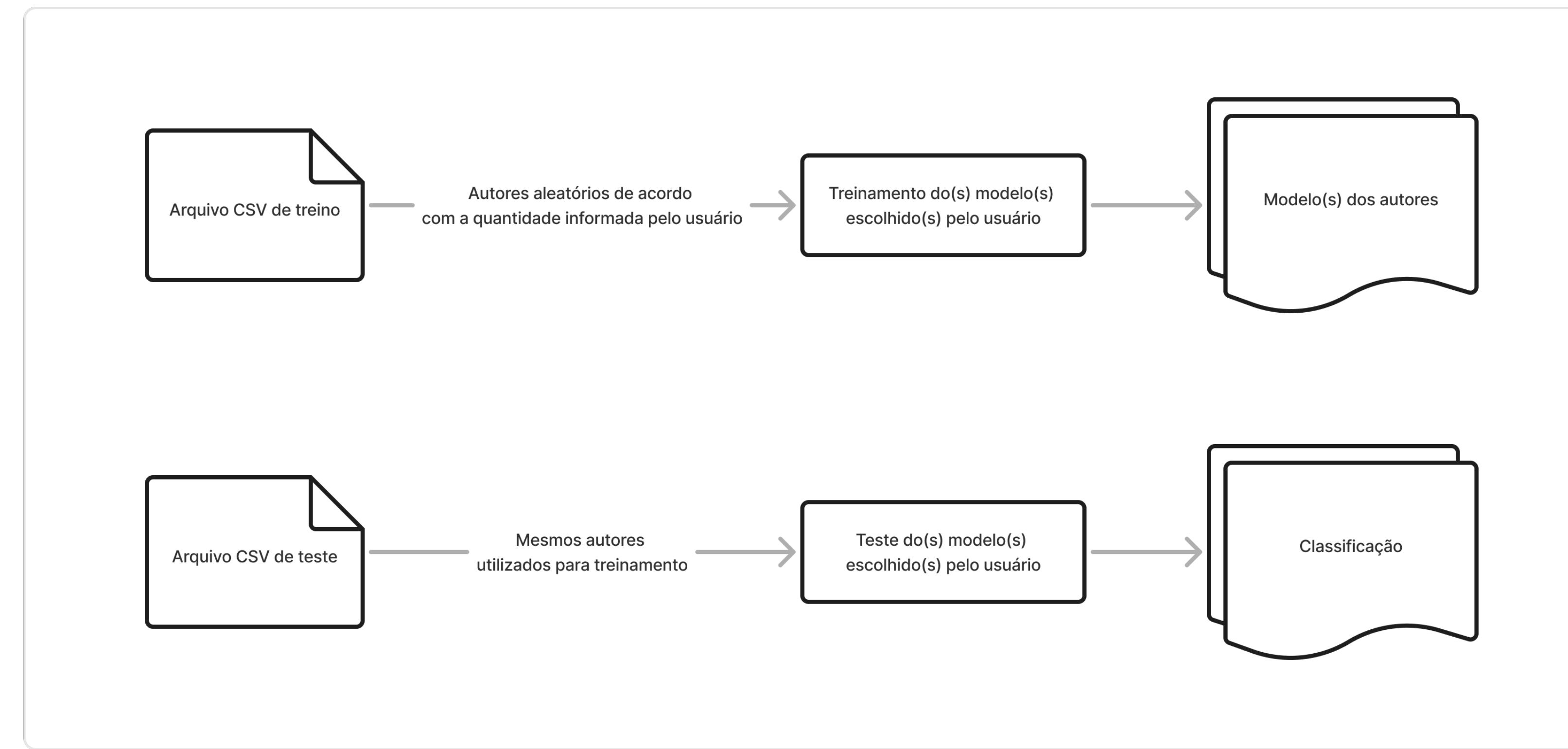


# Comunicação entre frontend e backend



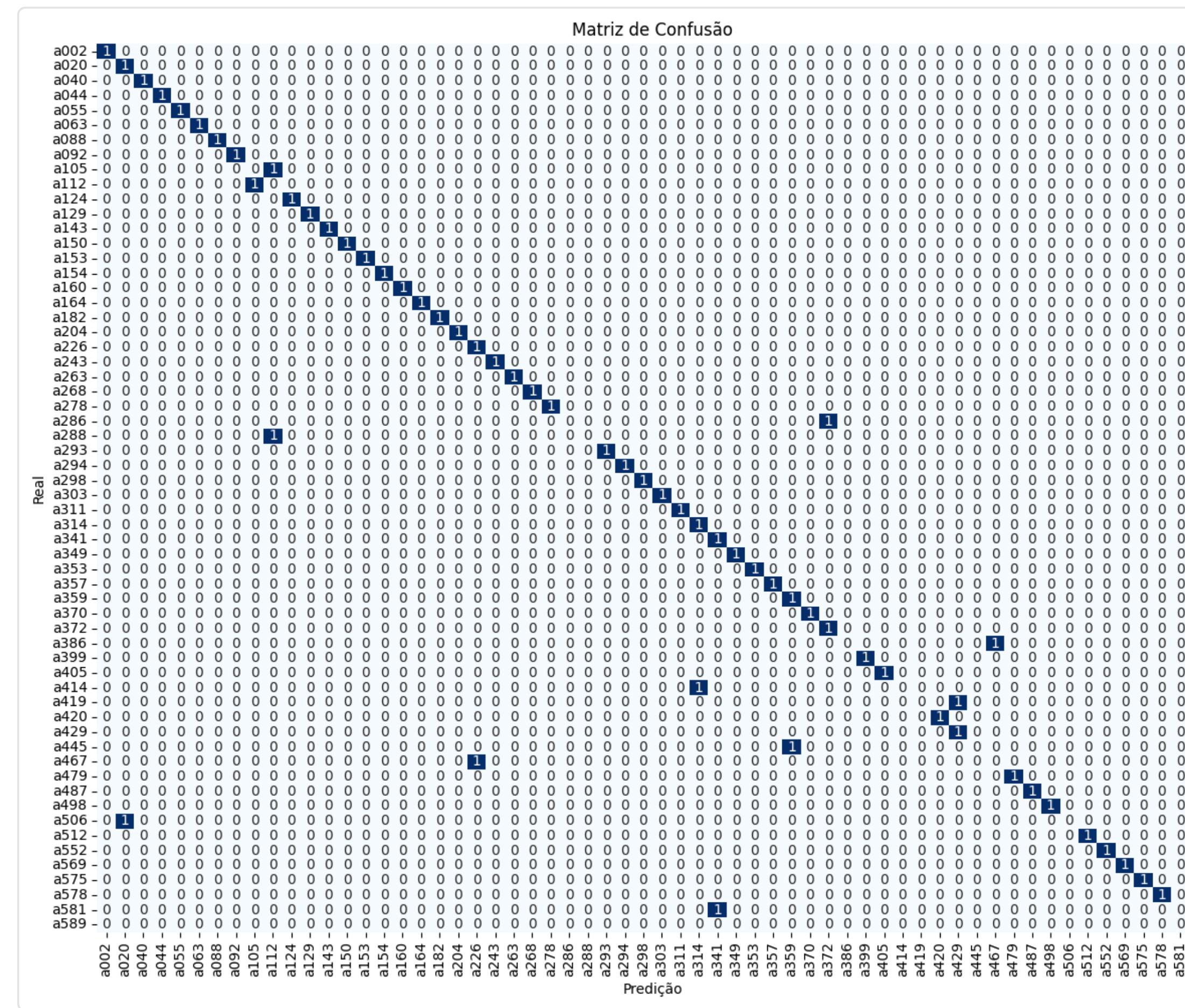
# Funcionamento do backend

Com os parâmetros recebidos, o backend carrega os dados de treino e teste dos arquivos CSV, seleciona aleatoriamente os autores de acordo com a quantidade especificada pelo usuário para compor os conjuntos de treino e teste e, finalmente, treina e testa os modelos selecionados pelo usuário, retornando as acurárias obtidas.



# Funcionamento do backend

Além da acurácia, o backend gera uma matriz de confusão para o modelo SVM, permitindo a visualização do desempenho do modelo em cada classe (autor) e a análise dos erros de classificação.



# **06 - Resultados**

# Melhores resultados

A fim de analisar a influência do número de autores no desempenho dos modelos e determinar a quantidade ideal de escritores para estabilizar os resultados, foi adotada uma metodologia similar à de Amaral (2014), incrementando o conjunto de dados em intervalos de 20 autores.

Número de Escritores	Taxa de Acerto (%)		
	SVM	SVM com Grid Search	Random Forest
20	95,00	95,00	90,00
40	80,00	92,50	82,50
60	78,33	81,67	78,33
80	68,75	76,25	76,25
100	67,00	71,00	68,00
120	70,83	76,66	71,66
140	59,28	69,28	62,14
160	61,87	71,25	65,00
180	56,11	67,77	67,22
200	55,00	65,50	59,50

## Análise e discussão dos resultados

Observando os resultados, destaca-se o desempenho superior do SVM com Grid Search em relação aos outros dois modelos, em todos os cenários testados.

Podemos dizer que o Grid Search, ao otimizar os hiperparâmetros do SVM, permitiu uma melhor adaptação do modelo aos dados, resultando em maior acurácia.

Os resultados sugerem que em cenários com um número reduzido de potenciais autores, a ferramenta apresenta alta eficiência na identificação de autoria.

À medida que o número de autores aumenta, observa-se uma queda gradual na acurácia para todos os modelos. Essa queda é esperada, uma vez que a complexidade do problema aumenta com o número de classes a serem classificadas.

# **07 - Conclusão**

# Conclusão

Os resultados obtidos demonstram a viabilidade da abordagem proposta, com o SVM otimizado por Grid Search apresentando o melhor desempenho dentre os classificadores avaliados.

A pesquisa contribui para a área de computação forense, oferecendo uma ferramenta que pode auxiliar peritos grafotécnicos na análise de manuscritos, automatizando parte do processo e reduzindo a subjetividade inerente à análise manual.

Apesar dos resultados promissores, identificou-se a necessidade de expandir o conjunto de características grafométricas analisadas para alcançar um desempenho comparável ao de outros trabalhos na literatura, como o de Amaral (2014).

# Trabalhos futuros

outras características grafométricas

técnicas de pré-processamento mais robustas

utilização de outros algoritmos de aprendizado de máquina

Além disso, a expansão da base de dados com amostras de manuscritos que apresentem desafios específicos para a análise forense, como:

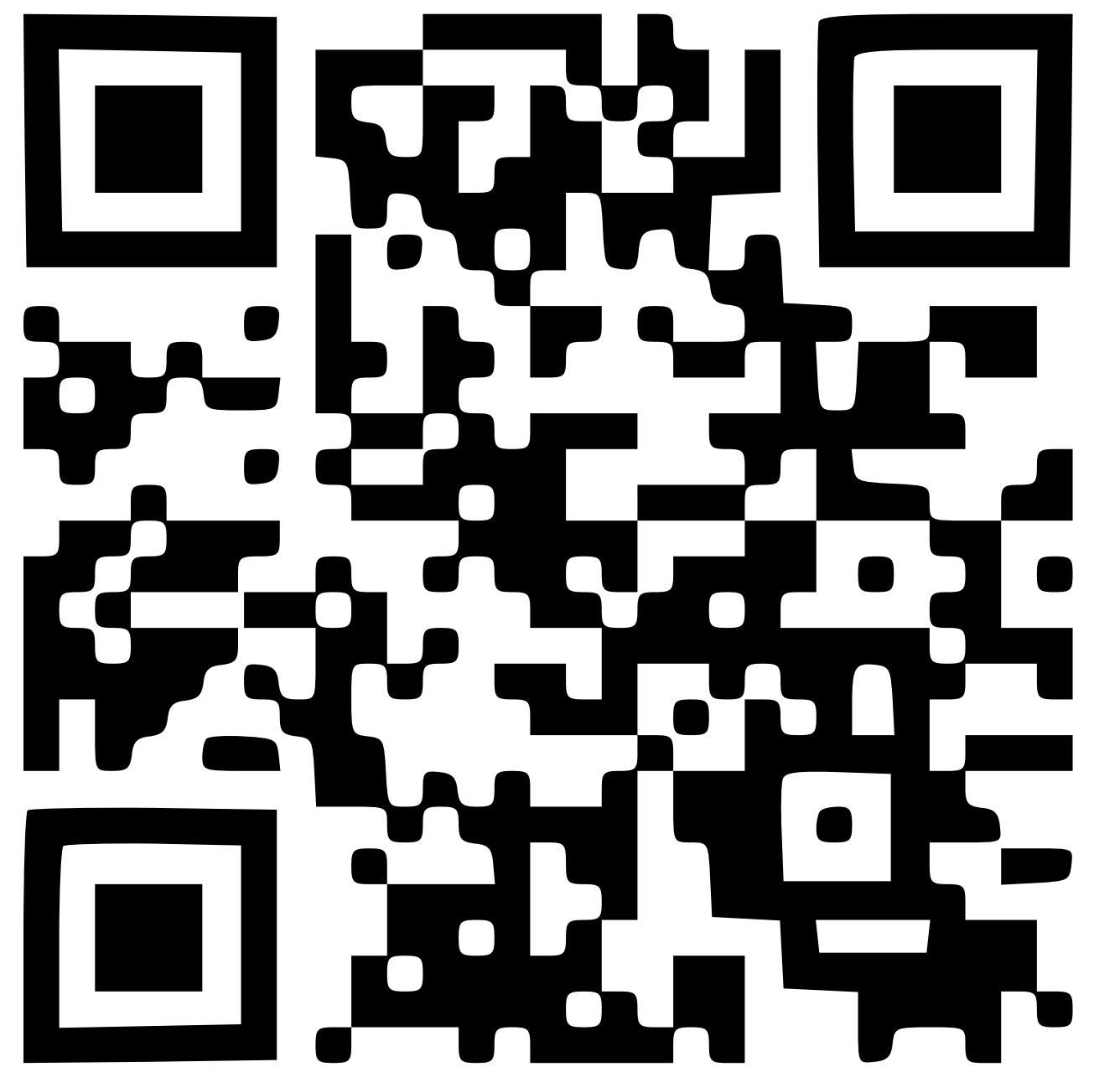
textos curtos

escritos com diferentes graus de legibilidade

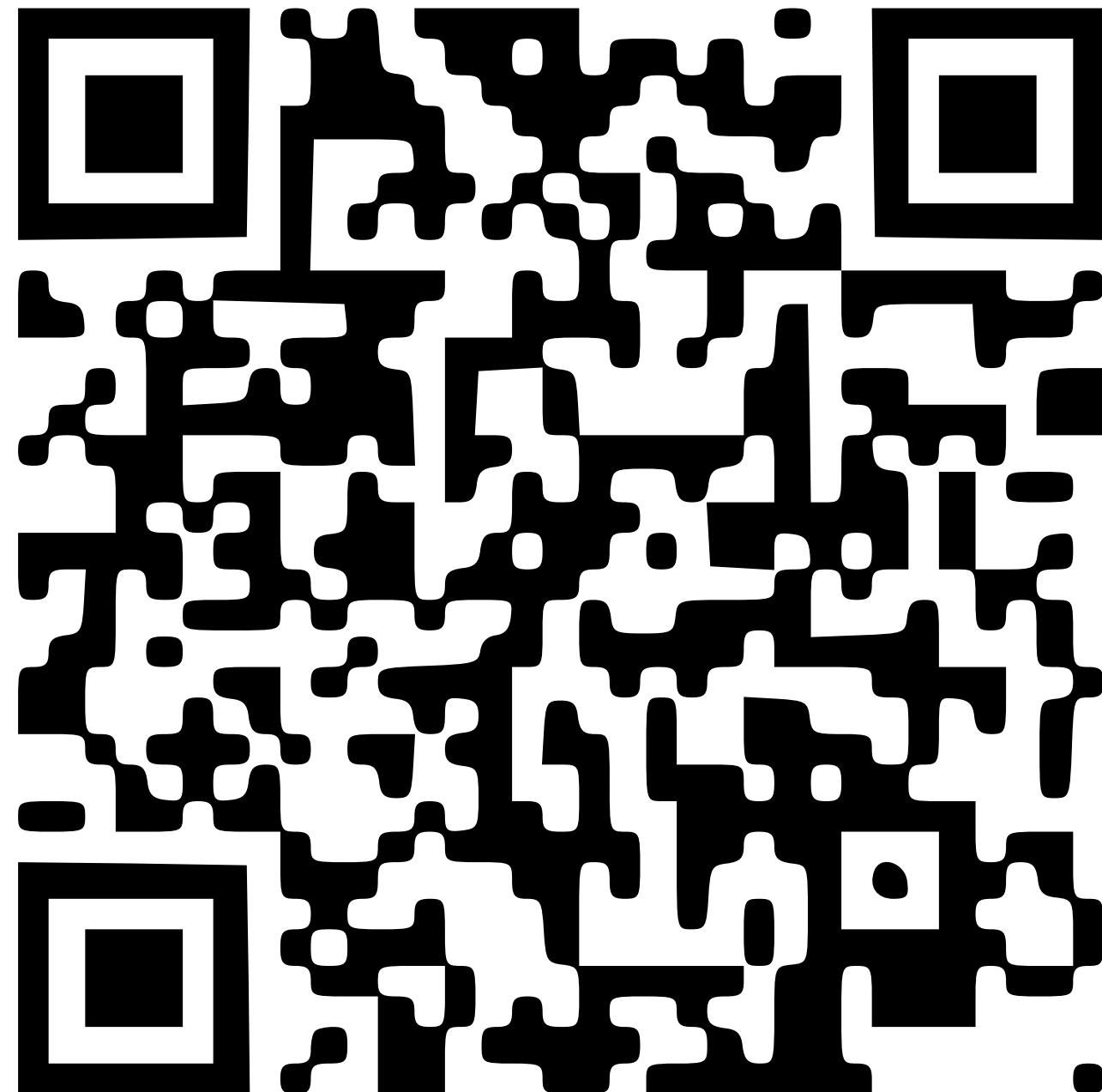
escritos disfarçados

avaliação do desempenho da ferramenta em cenários forenses reais

são importantes passos para a consolidação da ferramenta como uma solução prática para a identificação de autoria em manuscritos.



<https://manuscritus.vercel.app/>



<https://github.com/yuripiresalves/manuscritus>

**Obrigado!**