Detecção de Documentos Acadêmicos Falsificados: Uma Solução Baseada em Aprendizado de Máquina

Samuel M. Ransolin[1], **Giovana N. Inocêncio**[1], **Jean E. Martina**[1]

[1]Departamento de Informática e Estatística – Centro Tecnológico
Universidade Federal do Santa Catarina (UFSC) – Florianópolis, SC – Brasil

samuel.moreira.ransolin@grad.ufsc.br,
{giovana.inocencio,jean.martina}@ufsc.br

Abstract. *In recent years in Brazil, the growth in entrants, graduates, and higher education institutions has intensified challenges in validating academic credentials, since verification remains largely manual, error-prone, and vulnerable to fraud. This article revisits the state of the art in machine-learning-based detection of forged academic documents and proposes a hybrid prototype that combines multimodal analysis, clustering, anomaly detection, and graded classification to assign a legitimacy score. By integrating the prototype into Jornada do Estudante, documents can be automatically validated prior to recording on a distributed ledger, thereby enhancing the security and reliability of credential issuance.*

Resumo. *Nos últimos anos, no Brasil, o crescimento de ingressantes, de formandos e de instituições de ensino superior intensificou os desafios relacionados à validação de certificados acadêmicos, já que a verificação é majoritariamente manual, sujeita a erros e a aceitação de fraudes. Este trabalho revisita o estado-da-arte em detecção de documentos falsificados via aprendizado de máquina, e propõe um protótipo híbrido que combina análise multimodal, clustering, detecção de anomalias e classificação por grau de legitimidade. Ao integrar o protótipo à Jornada do Estudante, documentos podem ser validados automaticamente antes do registro em sua rede distribuída, aumentando a segurança e a confiabilidade do credenciamento.*

1. Problemática

Com o decorrer da última década, presencia-se no Brasil crescimento contínuo na emissão de diplomas de ensino superior, com cifras que chegam ao aumento de mais de 31% de formandos desde 2013 [BRASIL 2024]. Embora isso revele um saldo extremamente positivo, também traz à tona desafios que precisam ser superados, entre eles e temática explorada neste trabalho, a melhoria nos processos de regulação, supervisão e avaliação dessas emissões por parte do Ministério da Educação do Brasil (MEC).

Atualmente, a gerência, armazenamento e emissão de documentos acadêmicos, como diplomas e históricos escolares, é responsabilidade da instituição de ensino que os emite [MEC 1978]. Além disso, o processo, burocrático e não computadorizado, é suscetível a erros e até mesmo fraudes devido à ausência de transparência e redundância [Palma et al. 2019]. Assim, essa falta de modernização deixa brechas que são conhecidas e utilizadas por agentes mal-intencionados, possibilitando a criação de falsas instituições,

especializadas na venda de pacotes que incluem diversos certificados contrafeitos amparados em documentos oficiais adulterados, de forma a conferir aparência de legalidade a diplomas sem qualquer base acadêmica real [Dias and Leal 2022].

É neste cenário que o MEC, em parceria com o Ministério da Economia e diversas universidades federais, disponibiliza o sistema da Jornada do Estudante, que permite que discentes acompanhem suas trajetórias estudantis junto ao acesso a seus documentos acadêmicos pertinentes. Além disso, esse sistema também pode tornar-se uma plataforma conjunta para a emissão e registro destes certificados e até mesmo dados regulatórios das instituições de ensino superior [RNP 2023]. Em acordo a essa iniciativa, o presente trabalho trata da implementação e validação de um protótipo de software que combina diferentes técnicas de aprendizado de máquina, capaz de identificar certificados falsos antes de sua inserção neste ambiente.

2. Estado da Arte

A pesquisa acadêmica sobre identificação de documentos falsificados é escassa, especialmente quando comparada aos estudos sobre detecção de fraudes. A detecção de fraudes foca em adulterações de arquivos originais, como a mudança de notas, datas ou nomes, enquanto a de documentos falsificados busca identificar aqueles completamente forjados desde sua criação, sem terem sido emitidos por instituições oficiais, por exemplo. Essa distinção é importante porque a caracterização e o conjunto de desafios práticos diferem. Entretanto, os métodos e técnicas utilizadas muitas vezes se sobrepõem e complementam entre sí, como é o caso deste trabalho, que aproveita referências em ambas as áreas e busca acrescentar às poucas soluções encontradas para a classificação de documentos falsificados em sua concepção.

No domínio geral, predominam estratégias de visão computacional, como o artigo de [Jaiswal et al. 2022], que utiliza autoencoders convolucionais sobre imagens hiperespectrais para identificar incompatibilidades entre tintas, ou como o trabalho de [James et al. 2020], que introduziram outra perspectiva ao reformular o problema como comparação de grafos, em que obtém, via OCR, caixas delimitadoras de tamanho entre caracteres, utilizando-as para o treinamento de classificadores que detectam a manipulação de pixels. Alternativamente, também existem propostas, como a de [Boonkrong 2024], que utilizam funções hash e registros imutáveis, em blockchain, para verificação posterior.

As abordagens preventivas mais robustas combinam múltiplas tecnologias para melhorar a detecção, destacam-se: o trabalho de [Kim 2022], que integra blockchain e aprendizado de máquina para diplomas, onde hashes e máscaras geradas por redes (Mask R-CNN / Faster R-CNN) são registradas para verificação e consenso; o trabalho [Jain and Wigington 2019], que demonstra a eficácia da análise multimodal de características textuais e visuais, combinando extração OCR, representações textuais (ULMFiT, FastText, n-grams) e codificações visuais (VGG-16) com diferentes estratégias de fusão; e o trabalho de [Mohammed et al. 2024], que utiliza clustering sobre a extração de features visuais para a detecção de anomalias entre documentos.

3. Metodologia

A metodologia proposta tem como base uma abordagem de aprendizado não-supervisionado e detecção de anomalias através da análise e extração multimodal. Busca-se rotular documentos com base em um nível de probabilidade de fraudulência, para isso, utilizam-se extrações de características visuais (como textura, fonte, espaçamento, selos e assinaturas), textuais (como padrões linguísticos, formatação de números e distribuição de termos) e estruturais (como posição de campos, margens e tabelas), que, combinadas, permitem o agrupamento dos documentos em clusters que representam padrões dominantes normais. Em sequência, modelos de detecção de anomalias são utilizados para a criação de detectores de referência a partir dos clusters, possibilitando a classificação de um novo documento submetido, em tempo real, através da avaliação do grau de desvio em relação aos padrões aprendidos – quanto maior o desvio e escore de anomalia, maior a probabilidade de que o documento seja falsificado. Finalmente, essa pontuação é mapeada para categorias discretas de suspeita, fornecendo um nível de probabilidade de fraude para cada inserção.

A escolha dessa abordagem tem por base a premissa de que documentos falsificados apresentam inconsistências sutis, tornando-os atípicos em relação aos padrões estabelecidos por documentos legítimos, sendo detectáveis através da análise multimodal das características extraídas de diversos contextos. Assim, o processo completo consiste em duas etapas: treinamento dos modelos de referência e classificação de novos documentos.

3.1. Treinamento dos Modelos de Referência

A fase de treinamento inicia com a coleta de certificações acadêmicas diversas, seguida do pré-processamento através de técnicas de normalização de imagens e aplicação de OCR. Com o dataset formado, é realizada a extração multimodal de características visuais, textuais e estruturais dos documentos. Em sequência, com base nos dados obtidos na etapa anterior, é realizada a identificação de padrões utilizando algoritmos de clustering para identificar grupos de documentos com comportamentos similares, estabelecendo padrões dominantes de normalidade. Por fim, detectores de anomalias são treinados para cada padrão descoberto, gerando modelos de referência normais.

3.2. Classificação de Novos Documentos

O fluxo de classificação de um novo documento reutiliza o mesmo pipeline de pré-processamento e extração multimodal para garantir consistência na representação. O resultado é comparado contra todos os modelos de referência normal. Cada modelo calcula um escore de anomalia baseado na distância, ou similaridade, em relação ao padrão estabelecido. Essas pontuações representam, por fim, a probabilidade de falsificação do registro. Finalmente, utilizam-se métricas de consenso para categorizar o arquivo, isto é, classificá-lo como normal ou suspeito a partir de determinado limiar de pontos.

3.3. Extração Multimodal

Em mais detalhes, o módulo de extração multimodal, passo chave e utilizado tanto no fluxo de treino quanto no fluxo de classificação de uma submissão, permite aproveitar o mesmo pipeline de processamento capturando e unindo características independentes e, no contexto deste trabalho, complementares. Essa abordagem opera, em paralelo, três diferentes subprocessos de extração de features:

- Extração visual: extrai características ligadas ao layout, qualidade e consistência visual dos documentos. Inclui análise de textura, propriedades de fonte (espessura, tamanho, espaçamento), qualidade de assinaturas e selos, resolução de imagem, e padrões de cores e contrastes;
- Extração textual: utiliza modelos de processamento de linguagem natural para extrair características linguísticas e de formatação. Analisa padrões textuais, distribuição de termos, consistência na formatação de números, datas e códigos, além de verificar a coerência semântica do conteúdo;
- Extração estrutural: extrai características ligadas à organização espacial e estrutural dos documentos. Examina posicionamento de campos, formatação de tabelas, alinhamentos, margens, espaçamentos e a disposição geral dos elementos no documento.

Por fim, as características extraídas são normalizadas, submetidas a técnicas de redução dimensional e fundidas, o que resulta em uma representação completa, unificada e compacta de cada documento. Isso permite que o sistema detecte tanto fraudes grosseiras quanto inconsistências sutis presentes em contrafações bem elaboradas.

Referências

Boonkrong, S. (2024). Design of an academic document forgery detection system. *International Journal of Information Technology*, pages 1–13.

BRASIL (2024). Censo da educação superior 2023: notas estatísticas.

Dias, P. and Leal, A. (2022). Sites vendem diploma de curso superior para quem sequer pisou em sala de aula: 'documentação 100% original, emitida de dentro da universidade', diz atendente.

Jain, R. and Wigington, C. (2019). Multimodal document image classification. *2019 International Conference on Document Analysis and Recognition (ICDAR)*, pages 71–77.

Jaiswal, G., Sharma, A., and Yadav, S. (2022). Deep feature extraction for document forgery detection with convolutional autoencoders. *Computers & Electrical Engineering*, 99:107770.

James, H., Gupta, O., and Raviv, D. (2020). Ocr graph features for manipulation detection in documents.

Kim, S.-K. (2022). Blockchain smart contract to prevent forgery of degree certificates: Artificial intelligence consensus algorithm. *Electronics*, 11:2112.

MEC (1978). Portaria mec/dau n 33 de 2 de agosto de 1978.

Mohammed, S., Nwobodo, L., and Ekene, N. (2024). Certificate fraud verification model using clustered-based classification approach. *Exploremática Journal of Innovative Engineering and Technology*, 5(1):60–72.

Palma, L. M., Vigil, M. A. G., Pereira, F. L., and Martina, J. E. (2019). Blockchain and smart contracts for higher education registry in brazil. *International Journal of Network Management*, 29.

RNP, R. (2023). Blockchain da jornada acadêmica.