

UNIVERSIDADE FEDERAL DE VIÇOSA

Extração automática de informações em imagens de notas fiscais

Mateus Fonseca de Souza
Magister Scientiae

VIÇOSA - MINAS GERAIS
2025

MATEUS FONSECA DE SOUZA

Extração automática de informações em imagens de notas fiscais

Dissertação apresentada à Universidade Federal de Viçosa, como parte das exigências do Programa de Pós-Graduação em Ciência da Computação, para obtenção do título de *Magister Scientiae*.

Orientador: Alexandre Santos Brandao

**VIÇOSA - MINAS GERAIS
2025**

**Ficha catalográfica elaborada pela Biblioteca Central da Universidade
Federal de Viçosa - Campus Viçosa**

T

S729e Souza, Mateus Fonseca de, 1999-
2025 Extração automática de informações em imagens de notas
fiscais / Mateus Fonseca de Souza. – Viçosa, MG, 2025.
1 dissertação eletrônica (69 f.): il. (algumas color.).

Orientador: Alexandre Santos Brandão.
Dissertação (mestrado) - Universidade Federal de Viçosa,
Departamento de Informática, 2025.
Referências bibliográficas: f. 66-69.
DOI: <https://doi.org/10.47328/ufvbbt.2025.430>
Modo de acesso: World Wide Web.

1. Mineração de dados (Computação). 2. Processamento de
imagens - Técnicas digitais. 3. Cobrança de contas. 4. Serviços
de eletricidade. I. Brandão, Alexandre Santos, 1982-.
II. Universidade Federal de Viçosa. Departamento de
Informática. Programa de Pós-Graduação em Ciência da
Computação. III. Título.

CDD 22. ed. 006.312

MATEUS FONSECA DE SOUZA

Extração automática de informações em imagens de notas fiscais

Dissertação apresentada à Universidade Federal de Viçosa, como parte das exigências do Programa de Pós-Graduação em Ciência da Computação, para obtenção do título de *Magister Scientiae*.

APROVADA: 24 de março de 2025.

Assentimento:

Mateus Fonseca de Souza
Autor

Alexandre Santos Brandao
Orientador

Essa dissertação foi assinada digitalmente pelo autor em 31/10/2025 às 16:08:47 e pelo orientador em 31/10/2025 às 16:11:02. As assinaturas têm validade legal, conforme o disposto na Medida Provisória 2.200-2/2001 e na Resolução nº 37/2012 do CONARQ. Para conferir a autenticidade, acesse <https://siadoc.ufv.br/validar-documento>. No campo 'Código de registro', informe o código **ORXI.N6XN.3NMT** e clique no botão 'Validar documento'.

AGRADECIMENTOS

Este trabalho é resultado de uma caminhada compartilhada. Dessa forma, agradeço a todos que estiveram ao meu lado ao longo do mestrado.

Aos meus orientadores, Alexandre e Lucas, expresso minha gratidão pela orientação, apoio constante e pelas valiosas contribuições técnicas e acadêmicas. O conhecimento compartilhado e a confiança depositada foram fundamentais para o desenvolvimento deste projeto e para meu crescimento profissional.

Aos colegas de pesquisa e do grupo de trabalho, agradeço pela troca de experiências, colaboração e pela convivência ao longo desta jornada.

Agradeço à Universidade Federal de Viçosa pelas oportunidades de aprendizado e por proporcionar um ambiente acadêmico de excelência, essencial para o desenvolvimento deste trabalho.

À Fundação de Amparo à Pesquisa do Estado de Minas Gerais (FAPEMIG), agradeço pelo apoio, indispensável para a viabilização desta pesquisa. Registro também meu agradecimento à Digital Grid pela parceria e pelo suporte oferecido durante a realização deste trabalho. A colaboração foi fundamental para aproximar a pesquisa da aplicação prática e enriquecer a experiência acadêmica.

Por fim, minha mais profunda gratidão a toda minha família, em especial aos meus pais, Hélio e Lucimar, que sempre acreditaram em mim e foram minha base em todos os momentos. O amor, a dedicação e o apoio incondicional que recebi foram fundamentais para que eu pudesse enfrentar os desafios deste percurso com coragem e determinação. A eles, dedico cada conquista desta jornada. Sou imensamente grato por tudo que fizeram e continuam fazendo por mim.

Este trabalho foi realizado com o apoio das seguintes agências de pesquisa brasileiras: Coordenação de Aperfeiçoamento de Pessoal de Nível Superior – Brasil (CAPES) – Código de Financiamento 001, Fundação de Amparo à Pesquisa do Estado de Minas Gerais (FAPEMIG) e Conselho Nacional de Desenvolvimento Científico e Tecnológico (CNPq).

RESUMO

SOUZA, Mateus Fonseca de, M.Sc., Universidade Federal de Viçosa, março de 2025. **Extração automática de informações em imagens de notas fiscais.** Orientador: Alexandre Santos Brandao.

Este trabalho propõe uma metodologia para a extração automatizada de informações a partir de imagens de notas fiscais de energia elétrica, um processo essencial no setor elétrico, especialmente para o gerenciamento de crédito. A metodologia desenvolvida abrange diversas etapas de processamento de imagens e visão computacional. Inicialmente, é realizada a segmentação de instâncias, detecção do documento e correção de perspectiva por meio da transformação de perspectiva. Além disso, são propostos dois algoritmos para a correção de distorções de orientação; o primeiro utiliza a dilatação dos pixels de uma imagem binarizada, enquanto o segundo se baseia na Transformada de Fourier. Para melhorar a qualidade das imagens processadas, também é realizada a remoção de ruídos de alta frequência com um filtro gaussiano e realizado o realce de contornos com máscara de nitidez. A fim de garantir a qualidade das imagens utilizadas, é introduzido um método de análise de qualidade. Para tornar o processamento mais adaptável, são empregados classificadores de documentos baseados em imagens e textos. Além disso, é proposta a utilização do modelo de consultas do serviço Amazon Textract, treinado para compreender tanto o texto quanto o layout dos documentos, permitindo a extração automática de informações específicas das notas fiscais de energia elétrica. Em faturas da distribuidora CEMIG, o modelo treinado atingiu uma precisão de 0,973 no conjunto de testes. O sistema desenvolvido demonstrou ser capaz de extrair todas as informações necessárias com um tempo médio de processamento de 35,273 segundos. Dessa forma, o projeto apresenta uma solução eficaz para a extração automatizada de informações de notas fiscais de energia elétrica, com alto potencial para aplicações empresariais.

Palavras-chave: Documentos; Faturas; Extração de Informações; Visão Computacional; Energia Elétrica

ABSTRACT

SOUZA, Mateus Fonseca de, M.Sc., Universidade Federal de Viçosa, March, 2025. **Automatic extraction of information from invoice images.** Adviser: Alexandre Santos Brandao.

This work proposes a methodology for the automated extraction of information from images of electricity bills, a crucial process in the electric power sector, particularly for credit management. The proposed methodology encompasses multiple stages of image processing and computer vision. Initially, instance segmentation, document detection, and perspective correction through perspective transformation are performed. Furthermore, two algorithms are proposed for correcting orientation distortions: the first employs dilation on a binarized image, while the second is based on the Fourier Transform. To enhance the quality of processed images, high-frequency noise is removed using a Gaussian filter, and edge enhancement is applied through a sharpening mask. In order to ensure the quality of the images used, a quality analysis method is also introduced. To increase adaptability during processing, document classifiers based on both image and text data are employed. Additionally, the use of the query-based model from the Amazon Textract service is proposed. This model is trained to understand both the text and the layout of documents, allowing for the automatic extraction of specific information from electricity bills. On invoices from the utility provider CEMIG, the trained model achieved an accuracy of 0.973 on the test set. The developed system was able to extract all required information with an average processing time of 35.273 seconds. Thus, the project presents an effective solution for the automated extraction of information from electricity bills, with high potential for business applications.

Keywords: Documents; Invoices; Information Extraction; Computer Vision; Electric Energy

Lista de ilustrações

Figura 1 – Ilustração de geração compartilhada	10
Figura 2 – Casos tratados na etapa de processamento de imagens.	13
Figura 3 – Fluxograma do processamento da imagem.	14
Figura 4 – Fluxograma geral de processamento.	14
Figura 5 – Máscara prevista para o modelo e os vértices detectados.	17
Figura 6 – Identificação do contorno do documento.	18
Figura 7 – Identificação dos vértices do documento.	18
Figura 8 – Resultado da correção de perspectiva em exemplos das Figuras 2(b) e 2(c), respectivamente.	19
Figura 9 – Evolução das métricas durante o treinamento do modelo.	20
Figura 10 – Processo de remoção de ruído.	22
Figura 11 – O processo de dilatação dos pixels e de definição dos <i>bounding boxes</i> , representadas pelas caixas de textos detectadas.	25
Figura 12 – Distribuição dos ângulos obtidos no processamento usando <i>Z-Score</i>	26
Figura 13 – Aplicação da correção de perspectiva e orientação.	27
Figura 14 – Avaliação da imagem pela análise em frequência.	29
Figura 15 – Aplicação do algoritmo de correção de orientação - Exemplo 1. Imagem original à esquerda e corrigida à direita.	31
Figura 16 – Aplicação do algoritmo de correção de orientação - Exemplo 2. Imagem original à esquerda e corrigida à direita.	32
Figura 17 – Aplicação do algoritmo de correção de orientação - Exemplo 3. Imagem original à esquerda e corrigida à direita.	33
Figura 18 – Imagens usadas para as features de referência - Parte 1.	42
Figura 19 – Imagens usadas para as features de referência - Parte 2.	43
Figura 20 – Imagens usadas para as features de referência - Parte 3.	44
Figura 21 – Imagem usada para as features de referência - Parte 4.	45
Figura 22 – Comparação de nível de confiança entre imagem original e corrigida.	51
Figura 23 – Fluxograma de extração de informações.	54
Figura 24 – Fluxograma de processamento das respostas.	55
Figura 25 – Fluxograma de treinamento do adaptador	55
Figura 26 – Exemplo de métricas do modelo treinado fornecido pela interface	56
Figura 27 – Processo de rotulação dos dados	57

Lista de tabelas

Tabela 1 – Métricas de desempenho do modelo	21
Tabela 2 – Resultado detecção de orientação. Método de Dilatação dos Pixels . . .	30
Tabela 3 – Resultado detecção de orientação. Método de Espectro de Frequência. .	30
Tabela 4 – Matriz de confusão referente ao experimento 1, mostrando o desempenho da classificação de imagens com o modelo <i>EfficientNetB0</i>	46
Tabela 5 – Matriz de confusão referente ao experimento 2, mostrando o desempenho da classificação de imagens com o modelo <i>EfficientNetB0</i>	47
Tabela 6 – Matriz de confusão referente ao experimento 3, mostrando o desempenho da classificação de imagens com o modelo <i>EfficientNetB7</i>	48
Tabela 7 – Matriz de confusão referente ao experimento 4, mostrando o desempenho da classificação de imagens com o modelo <i>EfficientNetB7</i>	49
Tabela 8 – Matriz de confusão referente ao experimento 5, mostrando o desempenho da classificação por texto.	50
Tabela 9 – Número de faturas utilizados no treinamento em cada modelo.	58
Tabela 10 – Resultados de Precisão, Recall e F1-score para diferentes distribuidoras em cada respectivo conjunto de teste.	59
Tabela 11 – Comparação de dados de energia com adaptador treinado e sem treinamento.	61
Tabela 12 – Tempo médio de execução do sistema de leitura.	62

Sumário

1	Introdução	9
1.1	Tema e Justificativa	12
1.2	Objetivos	12
1.3	Definição do Problema e Abordagem Proposta	13
1.4	Estrutura da Dissertação	15
2	Detecção e Correção do Documento	16
2.1	Análise da Perspectiva	16
2.1.1	O Método de Correção Proposto	17
2.1.2	Resultados e Discussão	20
2.2	Remoção de Ruídos	21
2.3	Análise da Orientação	22
2.3.1	Detecção de Orientação Baseada na Dilatação dos Pixels	24
2.3.2	Detecção de Orientação Baseada no Espectro de Frequência	26
2.3.3	Resultados e Discussão	28
3	Classificação de Concessionárias pela Identificação Automática das Faturas	34
3.1	CNNs na Extração de Características e Classificação de Imagens	36
3.2	Classificação de documentos por imagem e por texto	38
3.2.1	O Desenvolvimento da Proposta	38
3.2.2	Resultados e Discussão	40
3.2.3	Consideração Finais sobre o Processo de Classificação	47
3.3	Qualidade da Imagem e o Impacto no Pré-processamento	49
3.3.1	Metodologia para Avaliação da Qualidade da Imagem	50
3.3.2	Resultados e Discussão	51
4	Extração de Informações dos Documentos	53
4.1	Metodologia de Extração de Informações	54
4.1.1	Treinamento dos Modelos	55
4.1.2	Validação e Pós-Processamento das Respostas	56
4.2	Resultados e Discussão	58
5	Considerações Finais	63
	Referências	66

1 Introdução

O sistema de compartilhamento de energia renovável introduz uma nova dinâmica no consumo e na distribuição de energia, permitindo que proprietários de sistemas de geração compartilhem o excedente de energia que produzem. Esse excedente pode ser distribuído para outros consumidores, que, por sua vez, podem utilizá-lo como crédito em suas próprias faturas de energia elétrica.

Desde 17 de abril de 2012, quando entrou em vigor a Resolução Normativa ANEEL nº 482/2012, o consumidor brasileiro pode gerar sua própria energia elétrica a partir de fontes renováveis ou cogeração qualificada. Pode até mesmo fornecer o excedente para a rede de distribuição de sua localidade, para posterior compensação do consumo de energia verificado. Trata-se da Microgeração e da Minigeração Distribuídas de Energia Elétrica – MMGD e do Sistema de Compensação de Energia Elétrica – SCEE, inovações que aliam economia financeira, consciência socioambiental e autossustentabilidade ([ANEEL, 2025](#)).

O Sistema de Compensação de Energia Elétrica (SCEE) permite que, quando a energia gerada em determinado mês for superior à quantidade consumida no mesmo período, o excedente possa ser distribuído a outras unidades consumidoras, de acordo com a modalidade de participação no SCEE, ou transformado em crédito para compensação do consumo nos meses seguintes ([ANEEL, 2025](#)). Dentre as modalidades de participação no SCEE, destacam-se:

Autoconsumo local: Ocorre quando a energia é gerada e compensada no mesmo local onde está instalada a Micro ou Minigeração Distribuída (MMGD).

Autoconsumo remoto: Permite que a energia seja gerada em um local e compensada em outro, desde que ambas as unidades consumidoras pertençam ao mesmo titular.

Geração distribuída em múltiplas unidades consumidoras: A energia gerada pode ser repartida entre os condôminos, seguindo porcentagens ou ordens de prioridade definidas pelos próprios consumidores.

Geração compartilhada: Possibilita que diversos interessados se unam por meio de consórcio, cooperativa, condomínio civil voluntário ou edifício, ou qualquer outra forma de associação civil instituída para esse fim, a fim de instalar uma ou mais centrais de MMGD e utilizar a energia gerada para compensação do consumo de todos os participantes.

A Figura 1 ilustra o sistema de compartilhamento através da geração compartilhada.

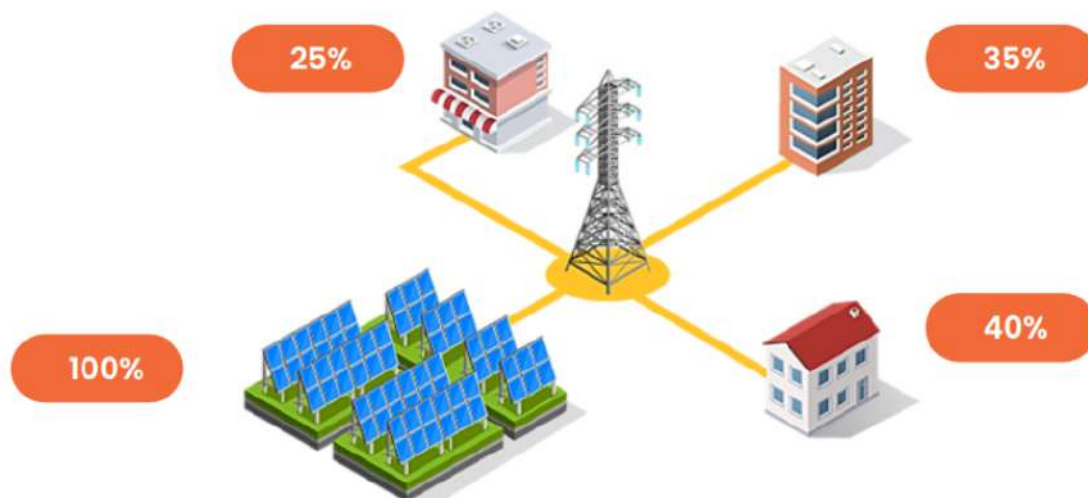


Figura 1 – Ilustração de geração compartilhada. (Projeto Solar, 2025)

A criação e administração de cooperativas de usinas solares possibilita que consumidores sem espaço próprio para instalação, como estabelecimentos comerciais em áreas urbanas e moradores de apartamentos, tenham acesso a geração distribuída. Um exemplo dessa tendência foi a implementação, em 2021, de uma plataforma digital em Minas Gerais para intermediar a assinatura de contratos voltados ao uso de créditos de compensação de geração distribuída (GD) solar (RIGO et al., 2022).

Nesse contexto, a leitura dos dados contidos na nota fiscal de energia elétrica permite determinar tanto a quantidade de energia gerada e injetada na rede elétrica quanto os créditos de energia disponibilizados para compartilhamento, além de fornecer diversas outras informações relevantes. Dessa forma, a análise desses dados possibilita a prevenção de eventuais erros ou desequilíbrios na distribuição do excedente de energia, assegurando que cada consumidor receba a quantidade correta de créditos correspondentes à energia compartilhada.

A gestão eficiente de crédito de energia através da extração de dados de notas fiscais pode ser uma tarefa desafiadora devido à sua diversidade de formatos, estilos e necessidade de extrair informações específicas. Um exemplo dessa proposta para gerenciamento eficiente dos créditos de energia é a implementação também em Minas Gerais de uma plataforma SaaS que simplifica a gestão de créditos das usinas, realizando a gestão completa dos créditos de energia. O funcionamento da plataforma tem como ponto de partida o download automático das faturas dos clientes finais e da usina de geração distribuída, seguido da extração de informações essenciais, como consumo, compensação e geração de energia. Com base nesses dados, o sistema possibilita a realização do faturamento da energia compensada e a otimização do cálculo da distribuição de créditos (DIGITAL GRID, 2025).

Embora o armazenamento de metadados seja uma tendência para minimizar a necessidade de processamento de documentos em papel, uma quantidade significativa de documentos ainda é impressa ou gerada no formato PDF (HA; HORÁK, 2022). Quando digitalizados, esses documentos podem estar sujeitos a condições adversas de captura, como iluminação inadequada, variações de perspectiva e orientação, fatores que podem comprometer significativamente a qualidade da imagem e, conseqüentemente, as análises subsequentes. Ademais, a digitalização pode resultar em imagens de baixa resolução, frequentemente afetadas por ruído, seja em decorrência da qualidade reduzida do equipamento utilizado, seja em razão das condições do próprio documento (SRIVASTAVA; HARIT, 2024).

Nesse contexto, o presente trabalho propõe contribuições que abrangem desde o processamento de imagens, com a remoção de ruídos e correção de orientação e perspectiva, até a leitura e extração de dados específicos a partir de imagens de documentos. Espera-se que, por meio dos métodos propostos, seja possível desenvolver uma ferramenta de extração de dados, capaz de processar em massa um grande conjunto de documentos. Em suma, este trabalho visa responder a seguinte pergunta:

Como um sistema automatizado de processamento de imagens e extração de dados pode melhorar a eficiência e a confiabilidade na gestão de créditos de energia?

Com base nessa questão, foram formuladas as seguintes hipóteses, que serão investigadas ao longo deste estudo:

- **H1:** A extração de dados de notas fiscais de energia elétrica pode ser automatizada com alta precisão por meio da aplicação de técnicas de Inteligência Artificial e Visão Computacional.
- **H2:** A melhoria na qualidade das imagens através dos métodos propostos, com a remoção de ruídos e correção de distorções, influencia positivamente na confiança da extração das informações das imagens.
- **H3:** A ferramenta automatizada proposta pode substituir o processamento manual, tornando-o mais eficiente em termos de tempo e recursos.
- **H4:** Os métodos desenvolvidos são generalizáveis para diferentes formatos e padrões de notas fiscais, permitindo sua aplicação em diversas distribuidoras de energia.

Essas hipóteses serão testadas por meio de experimentos e análises, visando validar a abordagem proposta.

1.1 Tema e Justificativa

Este trabalho tem como foco a extração automatizada de dados de notas fiscais de energia elétrica por meio de técnicas de Inteligência Artificial e Visão Computacional. O objetivo principal é desenvolver métodos capazes de identificar e extrair informações textuais contidas nesses documentos, otimizando a leitura e o gerenciamento de créditos de energia.

A relevância deste estudo reside na necessidade de automatizar o processamento de notas fiscais de energia elétrica, uma atividade que, quando realizada manualmente, demanda tempo e recursos consideráveis. Com a aplicação de Inteligência Artificial e Visão Computacional, busca-se desenvolver uma solução eficiente para extrair informações essenciais, como energia gerada, compensada e créditos acumulados, de maneira rápida e precisa. Além disso, a utilização dessas tecnologias viabiliza a conversão de documentos no formato PDF ou imagem em texto digital estruturado, permitindo análises automatizadas em larga escala.

Dessa forma, a automação desse processo não apenas otimiza recursos humanos e reduz o tempo necessário para a extração de dados, mas também amplia a aplicabilidade dessas soluções para outros contextos que envolvem grandes volumes de documentos, como faturas, contratos e recibos. Assim, este projeto contribui significativamente para a modernização e eficiência no processamento de informações, promovendo maior agilidade e confiabilidade na gestão de créditos de energia.

1.2 Objetivos

O objetivo geral deste trabalho é desenvolver uma ferramenta automatizada capaz de extrair informações em imagens de notas fiscais de energia elétrica de diversas distribuidoras, a fim de apoiar a gestão eficiente do compartilhamento de crédito de energia entre proprietários de sistemas de geração de energia renovável e outros consumidores.

No atual contexto, é possível destacar os seguintes objetivos específicos:

- Realizar a detecção automática das notas fiscais de energia elétrica, identificando a região do documento, de forma rápida e com alta confiança.
- Melhorar a qualidade das imagens das notas fiscais, removendo ruídos e deformações da geometria do documento para garantir que a região do documento seja clara e legível.
- Aplicar algoritmo para detecção e correção de orientação do documento.
- Implementar classificador das faturas dos documentos.

- Analisar e validar qualidade da imagem processada.
- Treinar modelo para extração de informações específicas dos textos presentes na imagem.
- Validar a precisão da ferramenta automatizada por meio de testes e avaliações, garantindo a acurácia na extração de informações das notas fiscais.

1.3 Definição do Problema e Abordagem Proposta

Neste trabalho, propõe-se a automação de todo o fluxo de processamento e extração de dados de notas fiscais de energia elétrica. O problema central consiste na necessidade de transformar imagens desses documentos em informações estruturadas de forma precisa e eficiente. Para isso, é essencial garantir que a imagem do documento seja aprimorada antes da extração automatizada dos dados.

Nesse contexto, aborda-se o processamento de uma imagem única do documento, onde a qualidade visual deve ser otimizada para permitir uma extração confiável das informações. Para tanto, considerando os desafios identificados na literatura e ao longo do desenvolvimento deste projeto, propõe-se um conjunto de técnicas para melhorar a qualidade e legibilidade das imagens digitalizadas. Essas técnicas incluem a detecção e correção de perspectiva do documento, a remoção de ruídos e a detecção e correção de sua orientação. A Figura 2 ilustra os principais casos tratados nesta etapa.

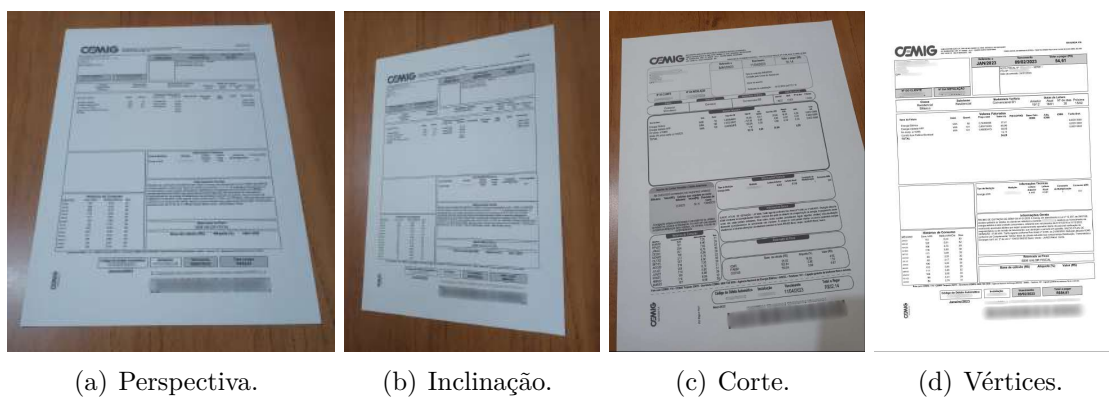


Figura 2 – Casos tratados na etapa de processamento de imagens.

Em particular, as Figuras 2(a) e 2(b) ilustram documentos com deformações causadas pela perspectiva, que modificam a projeção da geometria original no plano da imagem. Isso ocorre quando o documento, que originalmente possui formato retangular, perde suas proporções devido ao ângulo ou à forma como foi capturado. Já a Figura 2(c) apresenta um caso em que, além da perspectiva e da presença de plano de fundo, alguns dos vértices do documento não estão visíveis, tornando-o visivelmente incompleto. A Figura 2(d), por sua vez, apresenta um documento com rotação inadequada (orientação

incorreta) e vértices não visíveis, além de um plano de fundo com coloração semelhante à do documento, dificultando sua segmentação.

Para sistematizar o fluxo de processamento de imagens, a Figura 3 apresenta um fluxograma detalhando a ordem de aplicação dos métodos propostos, com a entrada (I) e a saída (O) representando, respectivamente, a imagem original e a imagem processada.



Figura 3 – Fluxograma do processamento da imagem.

Mesmo após o aprimoramento da imagem, é necessário avaliar sua qualidade para garantir que seja adequada à extração de dados. Assim, propõe-se uma análise específica para aferir a qualidade da imagem resultante. Além disso, no contexto do uso desses documentos em um sistema empresarial, é essencial identificar a distribuidora de energia responsável pela emissão da nota fiscal. Para isso, propõe-se um método de classificação automática das notas fiscais de acordo com a distribuidora.

Por fim, para realizar a extração das informações contidas nos documentos, é necessário um método robusto, seguro e com alta capacidade de generalização. O método adotado neste trabalho baseia-se em modelos de Respostas a Perguntas Visuais sobre Documentos, conhecidos como *Document Visual Question Answering (DVQA)*. A Figura 4 apresenta o fluxograma geral do processamento, detalhando todas as etapas envolvidas.

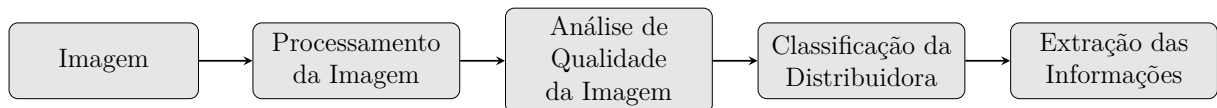


Figura 4 – Fluxograma geral de processamento.

É importante destacar que as imagens de notas fiscais de energia elétrica utilizadas neste projeto foram fornecidas pela *Digital Grid*, empresa parceira no desenvolvimento da presente pesquisa. Todos os experimentos foram conduzidos em um *Notebook* modelo *Latitude 3540*, fabricado pela *Dell Inc.*, operando com o sistema *Ubuntu 22.04.5 LTS*. A máquina está equipada com um processador *13th Gen Intel(R) Core(TM) i5-1335U*, 16 GB de memória RAM e unidade de armazenamento SSD de 256 GB. Nos capítulos seguintes, cada etapa de processamento das faturas ou das informações nelas contidas será conduzida de acordo com a fundamentação metodológica descrita e executada neste equipamento.

1.4 Estrutura da Dissertação

Esta dissertação está organizada em cinco capítulos, estruturados de forma a apresentar, de maneira progressiva e coerente, os desafios, métodos e resultados obtidos ao longo do estudo. O **Capítulo 1** introduz o contexto da pesquisa, destacando a relevância do processamento automatizado de imagens de documentos, além de apresentar a justificativa, os objetivos e a organização do trabalho. No **Capítulo 2**, são discutidas técnicas para a detecção e correção de documentos digitalizados, abrangendo a segmentação do documento na imagem, remoção de perspectiva e ajuste da orientação por meio de abordagens baseadas em Visão Computacional. O **Capítulo 3** trata da classificação das faturas de energia elétrica, utilizando redes neurais convolucionais para identificação automática das concessionárias de energia e abordando a análise da qualidade das imagens e seu impacto no desempenho do sistema. No **Capítulo 4**, apresenta-se a metodologia para extração automatizada de informações dos documentos, incluindo o treinamento de modelos, validação das respostas e pós-processamento dos dados extraídos, destacando a implementação do *Amazon Textract* e estratégias para aumentar a confiabilidade da extração. Por fim, o **Capítulo 5** sintetiza as considerações finais do estudo, discutindo as principais contribuições alcançadas, as limitações observadas e sugestões para trabalhos futuros, com foco na otimização do tempo de processamento e no aprimoramento da generalização dos modelos. Dessa forma, essa estrutura permite uma evolução lógica no desenvolvimento do trabalho, desde os desafios iniciais no processamento de documentos digitalizados até a implementação de soluções eficazes para extração e gerenciamento de informações relevantes.

2 Detecção e Correção do Documento

2.1 Análise da Perspectiva

A detecção e correção de perspectiva de documentos desempenham um papel fundamental no processamento automatizado de imagens, garantindo que os dados extraídos sejam precisos e confiáveis. Neste contexto, a detecção refere-se à localização e extração da região correspondente ao documento na imagem, bem como à sua classificação.

Considerando um cenário em que a extração de dados ocorre a partir de uma única foto digitalizada de um documento, a primeira etapa crítica no processamento é a identificação e separação do documento em relação ao plano de fundo ([ATTIVISSIMO et al., 2019](#)). A complexidade desse processo decorre de fatores como iluminação variável, ruídos na imagem e distorções introduzidas pela perspectiva no momento da captura.

Na literatura, diversos estudos abordam o problema da detecção e extração de documentos em imagens. [Attivissimo et al. \(2019\)](#) propõem um método para a extração de dados de documentos de identidade italianos, iniciando pela localização do documento na imagem, seguida de sua classificação. A avaliação do desempenho considerou a detecção correta dos vértices quando o erro máximo de deslocamento foi inferior a 3% do menor lado do documento, resultando em uma taxa de acerto de 68,57% na detecção e 100% na classificação.

Em outro estudo, [Tropin et al. \(2021\)](#) analisam a detecção de documentos por dispositivos móveis, propondo uma variação do método baseado em contornos, o que resultou em uma redução dos erros gerais de detecção em 10%. Já [Castelblanco et al. \(2020\)](#) exploram o uso de aprendizado de máquina para detecção de documentos enviados por usuários em serviços bancários. A proposta visa mitigar problemas durante a digitalização e obteve uma precisão de 98,4% na detecção do fundo da imagem.

No contexto específico da detecção de notas fiscais, [Sidhwa et al. \(2018\)](#) utilizam a biblioteca *OpenCV* para identificar e segmentar a fatura presente na imagem, além de aplicar técnicas de filtragem para remover ruídos. Além disso, os autores propõem a aplicação do *Tesseract OCR* para a extração de texto, aprimorando a eficiência do processo.

Dessa forma, os métodos propostos neste capítulo visam atender a demandas para a detecção e correção de documentos em imagens, detalhando estratégias para aprimorar a seleção, correção e estruturação das imagens, garantindo uma extração de dados mais confiável. A próxima seção apresenta os métodos adotados neste estudo.

2.1.1 O Método de Correção Proposto

Ao processar uma imagem de um documento, é essencial identificar e corrigir apenas a região correspondente ao documento, eliminando o plano de fundo. No entanto, em determinadas situações, além da remoção do fundo, o documento pode apresentar distorções de perspectiva. Essas distorções ocorrem devido ao ângulo de captura da imagem, tornando a extração de dados menos precisa. A Figura 2(a) ilustra um exemplo dessa problemática.

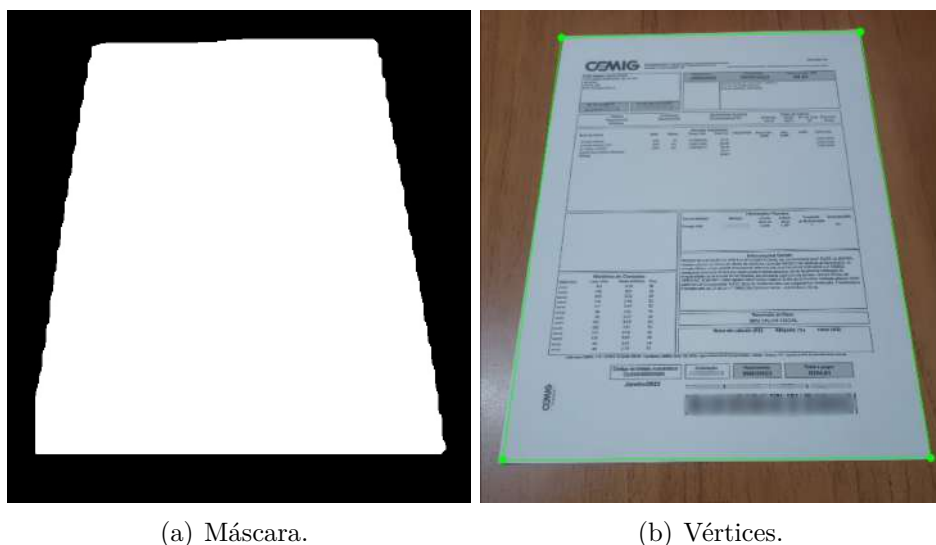


Figura 5 – Máscara prevista para o modelo e os vértices detectados.

Para lidar com esse desafio, propõe-se inicialmente a detecção do documento por meio de segmentação de instâncias. Esse método não apenas realiza a detecção do objeto, mas também a segmentação precisa de sua região na imagem. A segmentação de instâncias permite identificar a forma exata dos objetos, como descrito em [Jocher, Chaurasia e Qiu \(2023\)](#). Neste trabalho, utiliza-se a *YOLOv8* para essa tarefa. O treinamento do modelo parte da versão pré-treinada *yolov8s-seg.pt*, disponibilizada por [Jocher, Chaurasia e Qiu \(2023\)](#), empregando o conjunto de dados [Lung \(2024\)](#), que contém 1763 imagens, divididas em 1243 para treinamento, 348 para validação e 172 para testes. As imagens são rotuladas com a classe *document*, e o treinamento foi conduzido por 60 épocas, utilizando um lote de 32 imagens com dimensões de 640x480. A Figura 5(a) apresenta um exemplo de segmentação e previsão da máscara pelo modelo treinado.

Após a segmentação do documento, é necessário determinar seus quatro vértices para o cálculo da matriz de transformação de perspectiva. Neste caso, a identificação dos vértices inicia-se com a aplicação do algoritmo de detecção de contornos, disponibilizado pela biblioteca *OpenCV*. Esse processo é ilustrado na Figura 6.

Após detectar o maior contorno presente na imagem, extrai-se seu casco convexo, o qual representa a menor forma convexa que encapsula completamente o contorno original,

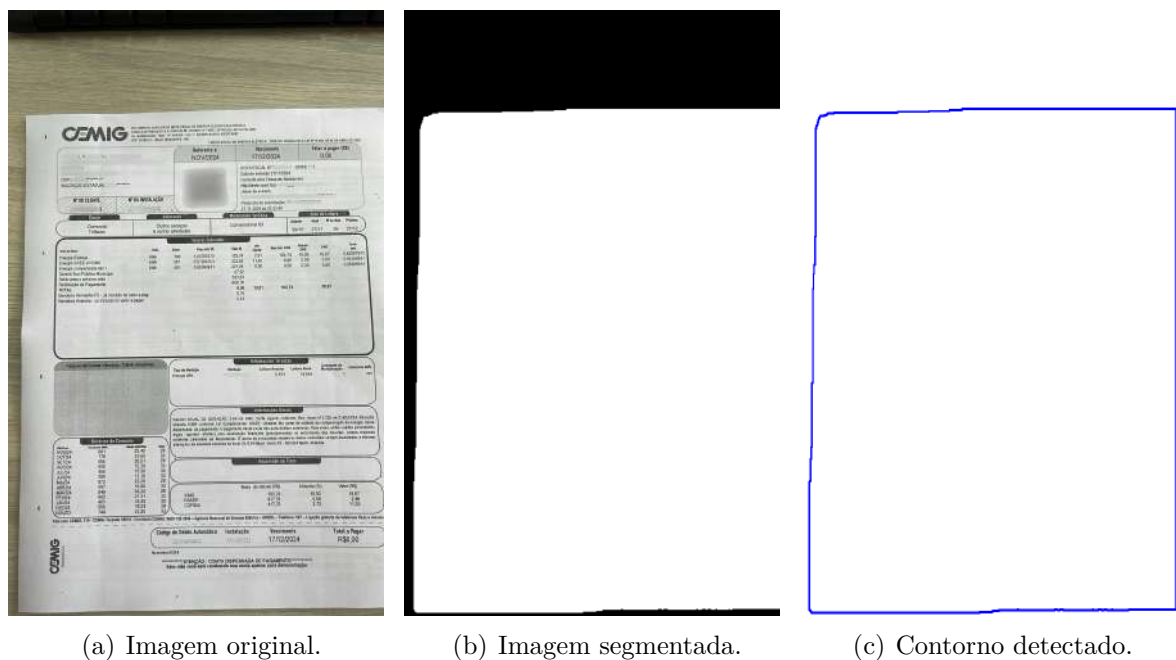


Figura 6 – Identificação do contorno do documento.

garantindo que todos os pontos do objeto estejam dentro da área delimitada. Esse passo é crucial, pois a borda da máscara segmentada nem sempre coincide exatamente com a borda do documento. A Figura 7(a) ilustra o casco convexo, evidenciando sua suavidade em relação ao contorno original (Figura 6(c)).

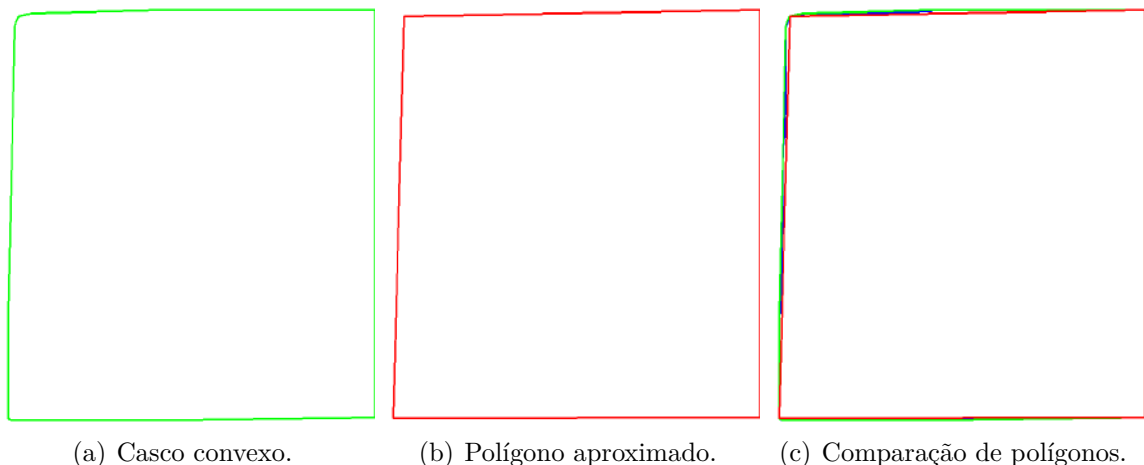


Figura 7 – Identificação dos vértices do documento.

Para garantir a correção adequada da perspectiva, realiza-se a aproximação do casco convexo a um polígono de quatro vértices. Esses vértices são então utilizados para calcular a matriz de transformação de perspectiva e aplicar a correção necessária. A Figura 7(b) apresenta um exemplo do polígono aproximado, enquanto a Figura 7(c) compara

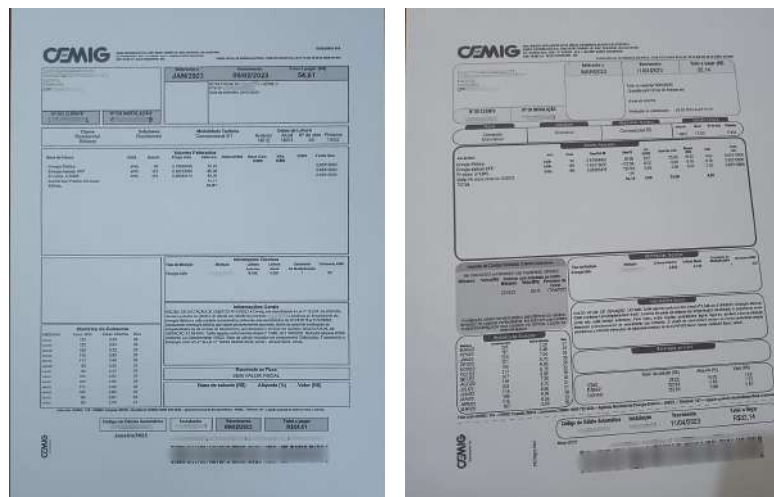
diferentes representações do contorno, mostrando como o casco convexo engloba a forma original e como a aproximação final resulta na definição dos quatro vértices.

Com os quatro vértices determinados, a matriz de transformação de perspectiva é calculada, e a correção é aplicada. O destino dos vértices é ajustado para um retângulo cujas dimensões são definidas pela maior diferença entre as coordenadas x e y dos quatro pontos detectados. A Figura 5(b) apresenta um exemplo da determinação dos vértices. Os métodos para extração do maior contorno, cálculo do casco convexo, aproximação de polígonos e transformação de perspectiva são fornecidos pela biblioteca *OpenCV*.

A extração de contornos é realizada através do método *cv2.findContours*. O casco convexo é obtido diretamente de cada contorno através de *cv2.convexHull*, que aplica o algoritmo de *Sklansky* com complexidade $O(N \log N)$ para detecção do conjunto de pontos. Para aproximação dos polígonos, emprega-se *cv2.approxPolyDP*, baseado no algoritmo de *Ramer–Douglas–Peucker* (BRADSKI, 2000).

Já a correção de perspectiva utiliza a função *cv2.getPerspectiveTransform* que recebe quatro pares de pontos correspondentes e resolve um sistema de equações homogêneas para gerar a matriz $M \in \mathbb{R}^{3 \times 3}$ que mapeia cada ponto de origem à sua contraparte no destino. Em conjunto é utilizado, *cv2.warpPerspective* aplicando essa matriz à imagem inteira, onde para cada pixel de saída (u, v) , calcula-se $(x, y) \sim M^{-1}[u, v, 1]^T$, preenchendo a imagem de destino com interpolação e modos de borda (BRADSKI, 2000).

Na Figura 8, observa-se o resultado da transformação aplicada a documentos com distorção de perspectiva.



(a) Correção Figura 2(b).

(b) Correção Figura 2(c).

Figura 8 – Resultado da correção de perspectiva em exemplos das Figuras 2(b) e 2(c), respectivamente.

2.1.2 Resultados e Discussão

Para avaliar o desempenho do modelo proposto para segmentação de documentos, são apresentados os resultados obtidos no conjunto de validação do *dataset* utilizado no treinamento. A Tabela 1 apresenta as principais métricas do modelo na tarefa de segmentação por instâncias, incluindo Precisão (P), *Recall* (R), *mAP50* e *mAP50-95*. Os resultados referem-se tanto à detecção de caixas delimitadoras quanto à detecção de máscaras, permitindo uma análise comparativa entre ambas as abordagens.

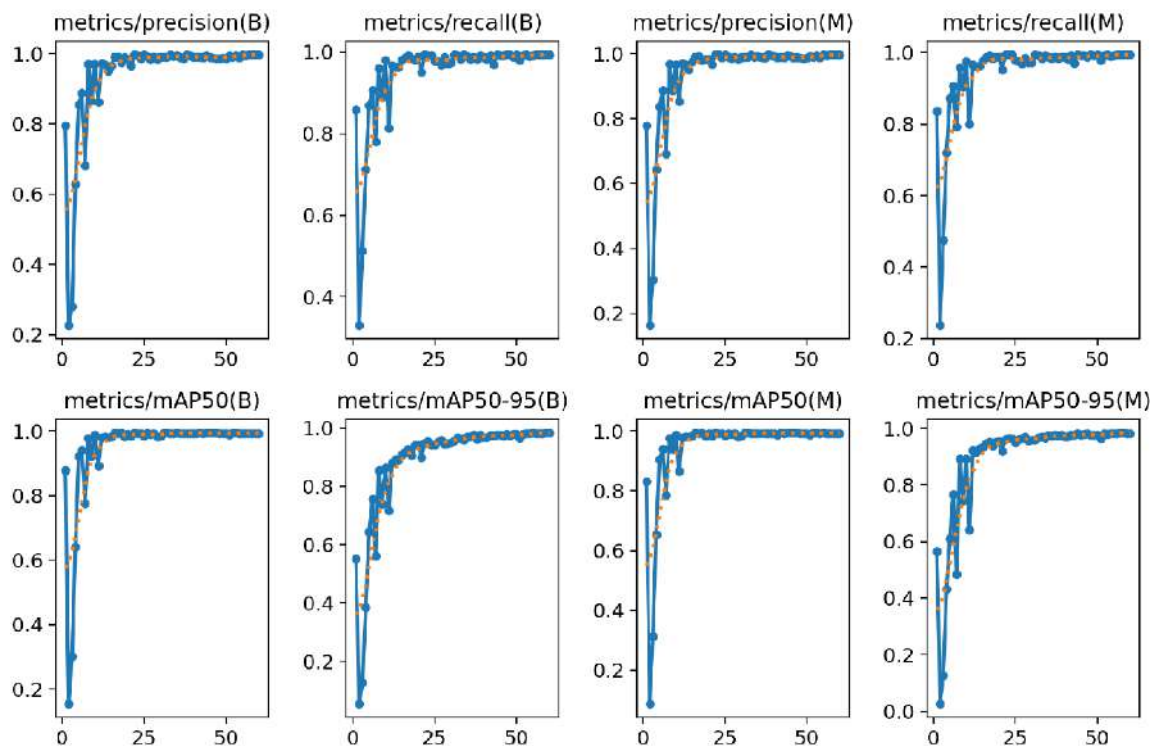


Figura 9 – Evolução das métricas durante o treinamento do modelo.

Na Figura 9, observa-se a evolução das métricas ao longo do treinamento do modelo. Nota-se que a estabilização dos valores ocorre após a época 25. Os resultados indicam um desempenho consistente entre as abordagens de detecção por caixas delimitadoras e por máscaras. Ambas alcançaram uma precisão de 0,997, *recall* de 0,994 e *mAP50* de 0,993. Entretanto, em relação à métrica *mAP50-95*, a detecção de máscaras apresentou um desempenho ligeiramente superior, atingindo 0,983, em comparação com 0,982 da detecção por caixas delimitadoras. Além disso, no conjunto de validação, o modelo falhou na segmentação em apenas duas instâncias, classificando incorretamente o documento como plano de fundo.

Com base nesses resultados, observa-se que o modelo treinado foi capaz de segmentar com alta precisão os documentos presentes nas imagens, atingindo uma taxa de acerto de 99,7%. Além disso, ao aplicar a técnica proposta para detecção de vértices e correção de perspectiva, foi possível remover completamente o plano de fundo e reduzir

Tabela 1 – Métricas de desempenho do modelo

	P	R	mAP50	mAP50-95
Caixas Delimitadoras	0,997	0,994	0,993	0,982
Máscaras	0,997	0,994	0,993	0,983

significativamente a distorção de perspectiva, tornando os documentos digitalizados mais alinhados geometricamente ao formato ideal para processamento automatizado.

2.2 Remoção de Ruídos

A correção da distorção de perspectiva é uma etapa fundamental para garantir uma base limpa e estruturada para o processamento subsequente. No entanto, ainda se faz necessário remover o ruído residual presente na imagem, pois pode comprometer a qualidade e a precisão das etapas posteriores, como a extração de textos. Assim, a remoção de ruídos é necessária para melhorar a confiabilidade e eficiência dos sistemas de OCR utilizados em análises de documentos (AHMAD; NAZ; RAZZAK, 2021).

Para abordar esse desafio, este trabalho propõe um método estruturado em três etapas principais. A primeira etapa consiste no aumento da resolução da imagem por meio da função `cv2.pyrUp`, disponibilizada pela biblioteca `OpenCV`. Essa função realiza uma ampliação piramidal da imagem, duplicando suas dimensões espaciais e, consequentemente, aumentando sua resolução (BRADSKI, 2000). Tal procedimento é aplicado a imagens consideradas de baixa resolução, com o intuito de evidenciar detalhes visuais sutis que poderão ser melhor aproveitados nas etapas subsequentes do processamento.

Na segunda etapa, aplica-se um filtro gaussiano utilizando a função `filters.gaussian`, fornecida pela biblioteca `Scikit-Image`. O objetivo desse filtro é suavizar a imagem e mitigar variações abruptas provenientes de ruídos. Para tanto, é adotado um valor de $\sigma = 0,2$, o que implica uma suavização leve, mantendo a integridade dos detalhes relevantes.

A terceira e última etapa consiste na aplicação de uma máscara de nitidez com o intuito de realçar os elementos textuais presentes na imagem, destacando-os em relação ao fundo e, assim, favorecendo sua posterior identificação e extração. Para isso, emprega-se a função `filters.unsharp_mask`, também disponibilizada pela biblioteca `Scikit-Image`. Nessa operação, o parâmetro `radius` é definido como 1, especificando o raio do desfoque utilizado na construção da máscara, enquanto o parâmetro `amount`, fixado em 1,5, determina a intensidade do realce aplicado à imagem.

O método proposto é descrito no Algoritmo 1. A Figura 10, exemplifica o processo aplicado a uma nota fiscal ampliada, demonstrando os resultados obtidos na eliminação de ruídos e no realce dos caracteres, evidenciando a eficácia do método proposto.

Valores Faturados		Valores Faturados	
Quant.	Preço Unit. R\$	Quant.	Preço Unit. R\$
30	0,83394409	30	0,83394409
195	0,65313000	195	0,65313000
195	0,68383415	195	0,68383415

(a) Imagem original.

(b) Imagem corrigida.

Figura 10 – Processo de remoção de ruído.

Algoritmo 1 Remoção de Ruído com Realce de Nitidez**Entrada:** Imagem original I **Saída:** Imagem processada I_f

- 1: $I \leftarrow \text{pyrUp}(I)$ {Aumentar resolução}
- 2: $I_s \leftarrow \text{GaussianFilter}(I, \sigma = 0.2)$ {Suavização leve}
- 3: $I_f \leftarrow \text{UnsharpMask}(I_s, \text{raio} = 1, \text{intensidade} = 1.5)$ {Aplicar máscara de nitidez}
- 4: **return** I_f

É importante ressaltar que o enfoque deste trabalho é desenvolver um método generalista, ou seja, capaz de lidar com diversas imagens sem a necessidade de ajustes específicos para cada caso. Essa abordagem busca evitar a especialização do algoritmo, que poderia comprometer seu desempenho em contextos variados, garantindo, assim, maior versatilidade na aplicação prática.

2.3 Análise da Orientação

A orientação de documentos digitalizados pode impactar no desempenho de sistemas de reconhecimento de texto, especialmente na presença de rotações significativas, sendo uma das etapas no pré-processamento de sistemas de reconhecimento de texto (RAMTEKE; PATHAN; MEHROTRA, 2011).

Diversos estudos na literatura abordam estratégias para a detecção e correção de inclinação em documentos. Al-Shatnawi e Omar (2009) desenvolvem um método baseado no centro de gravidade, no qual a orientação do documento é determinada ao inscrevê-lo em um polígono arbitrário e identificar sua linha de base a partir do ponto central. Segundo os autores, essa abordagem alcançou uma precisão de 87%. Já Singh, Bhatia e Kaur (2008), utilizam um processo dividido em três etapas: primeiro, aplicam um pré-processamento

baseado em *block adjacency graph* (*BAG*); em seguida, realizam a Transformada de *Hough*; por fim, corrigem a inclinação da imagem por meio de rotação.

Além das abordagens tradicionais, métodos baseados em aprendizado profundo também têm sido explorados para a detecção da orientação de documentos. [Akhter e Rege \(2020\)](#) propõem um modelo de aprendizado profundo para identificar o ângulo do documento, obtendo uma precisão de 99,06% no conjunto de dados de teste e 98,75% no conjunto de dados de referência. No entanto, a previsão do ângulo foi limitada a incrementos de 1 grau. Já [Boiangiu et al. \(2020\)](#) adotam uma abordagem inovadora, combinando diferentes algoritmos para a detecção de orientação, tornando o método aplicável a diferentes tipos de documentos, independentemente do seu formato ou estrutura.

No entanto, para ser eficaz, a detecção e correção de orientação deve apresentar características como rapidez, confiabilidade, resistência a ruídos e capacidade de operar dentro de uma faixa aceitável de ângulos ([AHMAD; NAZ; RAZZAK, 2021](#)).

[Ahmad, Naz e Razzak \(2021\)](#) argumentam que a inclinação durante a digitalização de documentos é inevitável e, para mitigar esse problema, propõem um método que utiliza a Transformada de *Hough* Probabilística para detectar linhas na imagem. Em seguida, essas linhas são agrupadas com base no paralelismo, sendo que o agrupamento com o maior número de linhas paralelas representa a inclinação predominante do documento.

[Abbas et al. \(2022\)](#) propõem um sistema de reconhecimento de documentos manuscritos que inclui a correção de inclinação como etapa essencial do processamento de imagem. Para isso, os autores modificam o algoritmo *Skew Detection using Center of Gravity*, melhorando sua eficiência na detecção da direção do texto.

No contexto da análise de imagens de documentos, a correção da orientação é uma etapa do processamento de imagens, contribuindo diretamente para a automação do reconhecimento e extração de informações. Entretanto, em determinadas situações, a detecção e correção da perspectiva podem ser comprometidas pela ausência de distinção clara entre o documento e o plano de fundo ou pela falta de visibilidade de seus quatro vértices. Esse problema pode ser observado nas Figuras 2(c) e 2(d), onde a falta de referências geométricas impede que o algoritmo de correção de perspectiva identifique corretamente os vértices e, conseqüentemente, realize a normalização da orientação.

A Figura 8(b) ilustra a aplicação do algoritmo de correção de perspectiva a imagem apresentada na Figura 2(c). Embora o fundo tenha sido removido, ainda persiste uma distorção na orientação do documento. Para solucionar essa limitação, propõem-se dois métodos distintos para a detecção da orientação, seguidos da correção da inclinação por meio da rotação da imagem. O primeiro método é a detecção de orientação baseada na dilatação dos pixels, enquanto o segundo método realiza a detecção de orientação baseado no espectro de frequência.

Esses métodos são comparados entre si para avaliar suas vantagens e desvantagens, visando determinar a estratégia mais eficaz para diferentes cenários de digitalização de documentos.

2.3.1 Detecção de Orientação Baseada na Dilatação dos Pixels

O método proposto fundamenta-se no processo de dilatação de pixels em uma imagem binarizada. Com essa operação, os caracteres individuais se unem, formando agrupamentos que correspondem a palavras e frases, gerando assim regiões de interesse. Esse processo pode ser visualizado na Figura 11(a).

Após a dilatação, todos os contornos presentes na imagem são identificados e filtrados com base na área ocupada. Regiões muito pequenas ou excessivamente grandes são descartadas, garantindo que apenas contornos representativos das palavras ou frases sejam mantidos. Em seguida, determina-se o menor retângulo que envolve essas regiões, permitindo a extração de caixas delimitadoras orientadas contendo as expressões textuais presentes na imagem. Um exemplo dessa detecção pode ser observado na Figura 11(b).

Com as caixas de texto extraídas, a inclinação do documento é determinada a partir da orientação dos lados superior e inferior de cada caixa em relação ao eixo horizontal. Assim, obtém-se um conjunto de ângulos que representam a inclinação dos textos na imagem.

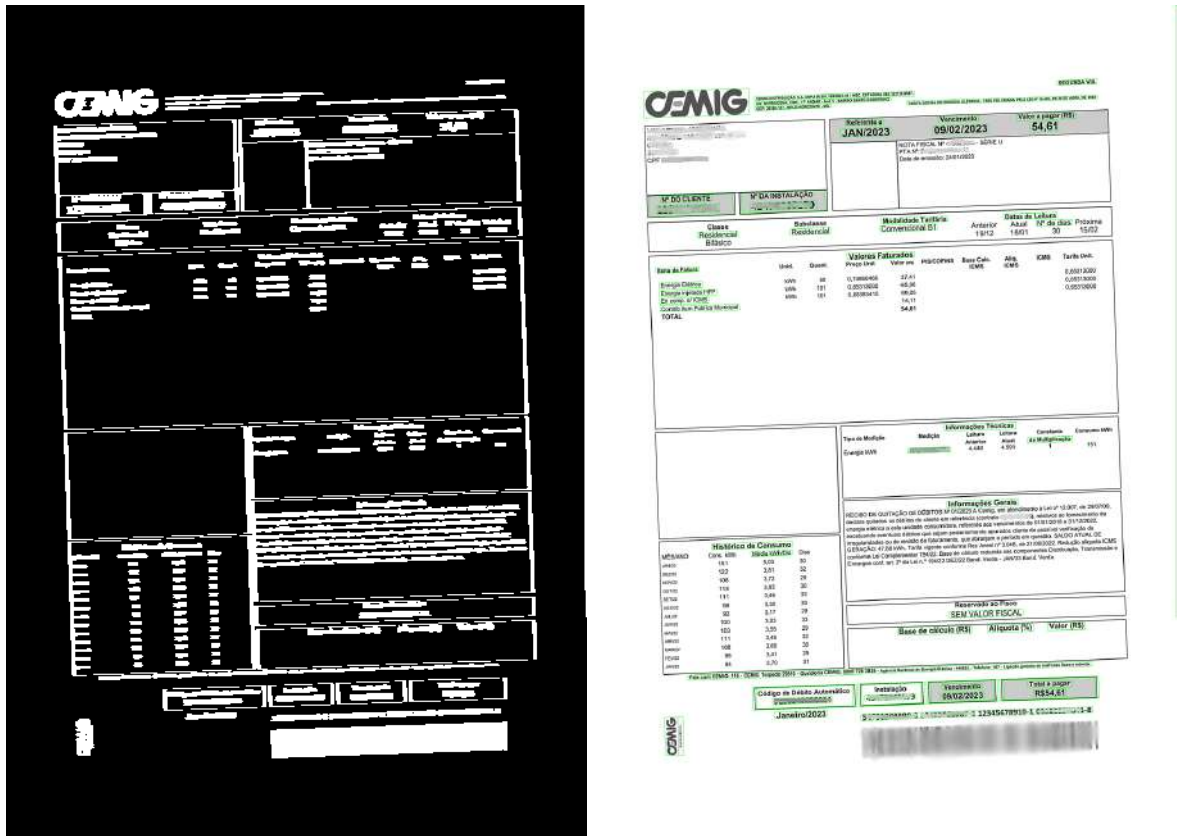
Para garantir que a correção da inclinação beneficie a maior parte do conteúdo textual presente na imagem, determina-se o ângulo predominante utilizando o método *Z-Score*. Esse procedimento permite filtrar os valores obtidos, removendo ruídos e assegurando que apenas ângulos relevantes contribuam para o cálculo da orientação final do documento.

O valor do *Z-Score*, denotado por Z , é definido por

$$Z = \frac{X - \mu}{\sigma}, \quad (2.1)$$

onde X representa um valor individual da amostra, μ é a média das amostras e σ corresponde ao desvio padrão. Com base nessa definição, as amostras são analisadas e aquelas cujo valor absoluto de Z excede um determinado limiar são excluídas do conjunto de ângulos. Para este experimento, foi adotado um limiar igual a 1, ou seja, qualquer amostra cuja diferença em relação à média do conjunto de ângulos ultrapasse o valor do desvio padrão é removida da análise.

A Figura 12 apresenta a distribuição dos ângulos extraídos das caixas de texto ilustradas na Figura 11(b), destacando tanto o ângulo predominante sem a aplicação do método *Z-Score* quanto sua evolução após a remoção dos valores discrepantes. Nota-se que, ao eliminar amostras extremas, o ângulo predominante ajusta-se melhor à orientação



(a) Imagem dilatada.

(b) Caixas de textos.

Figura 11 – O processo de dilatação dos pixels e de definição dos *bounding boxes*, representadas pelas caixas de textos detectadas.

esperada, reduzindo a influência de valores atípicos e aprimorando a precisão da correção de inclinação do documento.

Finalmente, após a filtragem dos ângulos, a orientação predominante é determinada por

$$\theta_{\text{pred}} = \frac{\sum_{i=1}^n A_i \cdot \theta_i}{\sum_{i=1}^n A_i}, \quad (2.2)$$

onde θ_{pred} representa o ângulo predominante, calculado como a média ponderada dos ângulos θ_i , considerando a área correspondente A_i de cada contorno identificado. Após a determinação do ângulo predominante, a imagem é rotacionada no sentido contrário para corrigir a inclinação.

Para exemplificar esta correção, a Figura 13 ilustra o resultado da aplicação do algoritmo de detecção e correção de perspectiva, seguido da correção da inclinação, onde no Caso 3, observa-se uma aproximação significativa com a imagem de um documento ideal, onde o fundo foi completamente removido e mesmo com vértices e bordas não visíveis, uma parte considerável da distorção de perspectiva foi corrigida. Além disso, a correção da orientação da imagem favoreceu a legibilidade da maioria das palavras, de modo que,

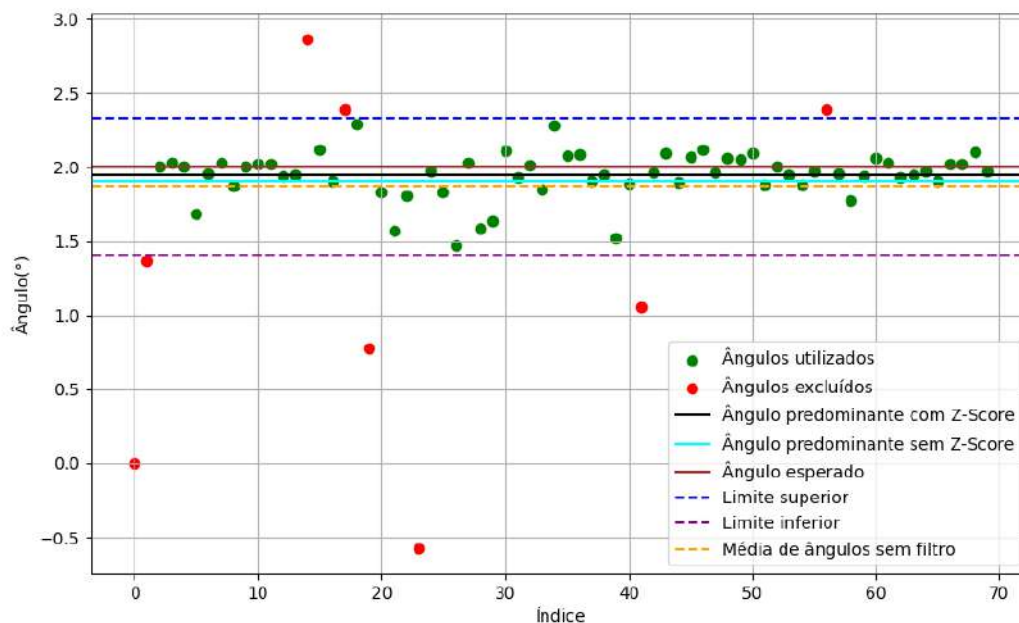


Figura 12 – Distribuição dos ângulos obtidos no processamento usando *Z-Score*.

embora persista alguma distorção, a maior parte do conteúdo textual é beneficiada pela correção aplicada.

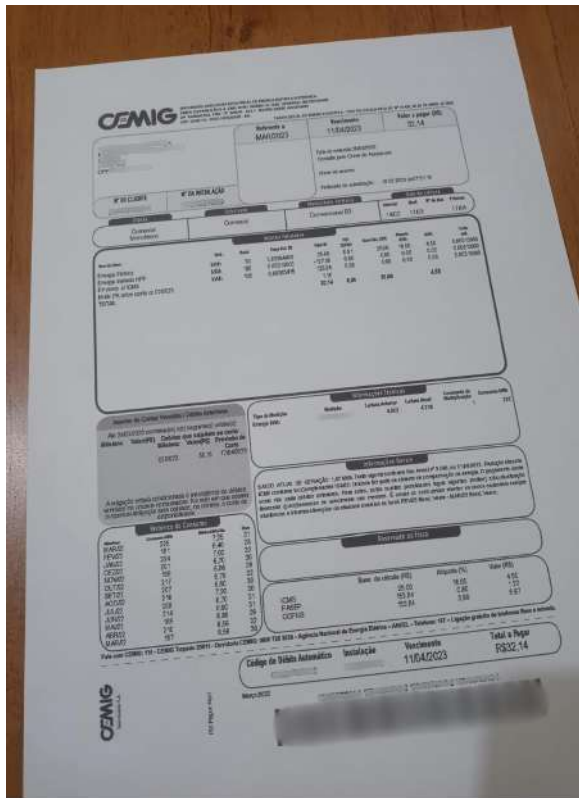
Por outro lado, no Caso 4, o sistema não é capaz de detectar nenhum vértice devido à semelhança do fundo com o documento original. No entanto, neste cenário, o sistema de correção de orientação consegue corrigir a distorção, ajustando a orientação do documento, mesmo sem a presença visível dos vértices. Isso demonstra a robustez do método na compensação de distorções de perspectiva e orientação, independentemente da visibilidade dos pontos de referência.

2.3.2 Detecção de Orientação Baseada no Espectro de Frequência

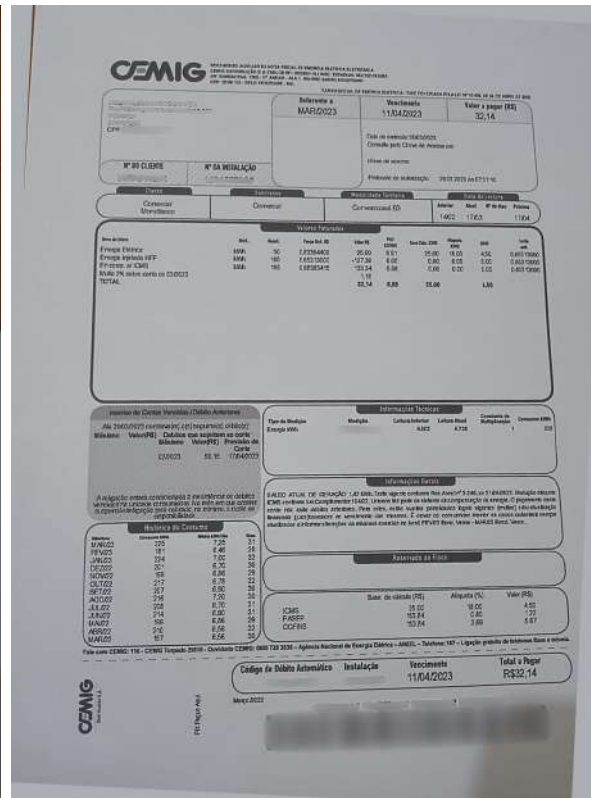
Uma abordagem alternativa para a correção de orientação de documentos consiste na análise do espectro de frequência da Transformada Discreta de Fourier. Esse método permite identificar e corrigir automaticamente o ângulo de inclinação da imagem, explorando padrões no domínio da frequência, como uma alternativa aos cálculos realizados no domínio espacial. A Figura 14 apresenta uma comparação entre a imagem no domínio espacial e no domínio da frequência.

Inicialmente, a imagem do documento em escala de cinza é submetida à DFT utilizando a função *cv2.dft*, da biblioteca *OpenCV*. Em seguida, é utilizado o método *np.fft.fftshift*, da biblioteca *NumPy*, para centralizar as baixas frequências no espectro.

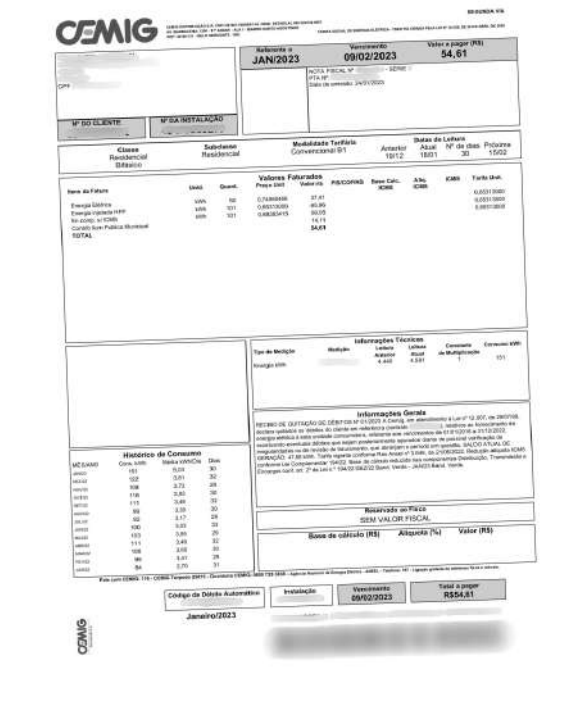
A magnitude do espectro foi então calculada por meio da função *cv2.magnitude*, e



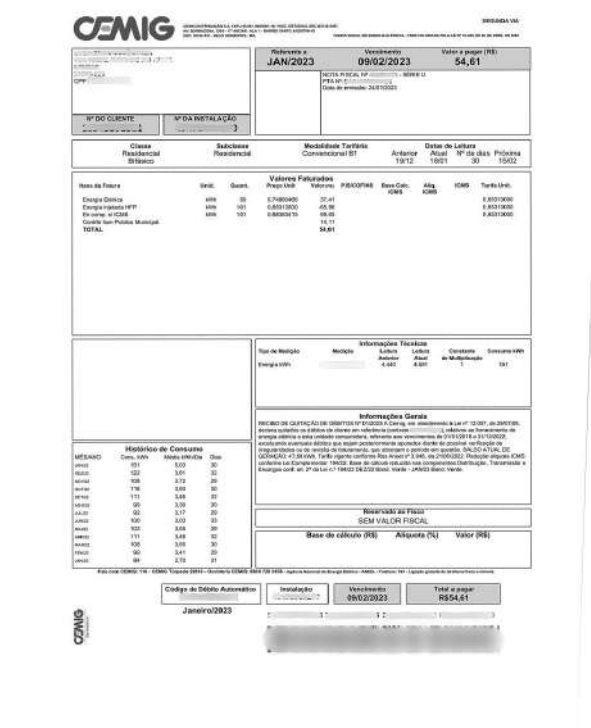
(a) Caso 3 original (Figura 2(c)).



(b) Caso 3 corrigido.



(c) Caso 4 original (Figura 2(d)).



(d) Caso 4 corrigido.

Figura 13 – Aplicação da correção de perspectiva e orientação.

posteriormente transformada em escala logarítmica com o objetivo de melhorar a visualização das componentes de alta frequência. Essa magnitude foi normalizada com *cv2.normalize*. A imagem resultante é redimensionada para 720×720 pixels utilizando *cv2.resize*, suavizada com um filtro Gaussiano via *cv2.GaussianBlur*, e segmentada com um limiar adaptativo (*cv2.adaptiveThreshold*) para destacar estruturas relevantes.

Após a segmentação, aplicou-se o operador de detecção de linhas por transformada de Hough probabilística, por meio da função *cv2.HoughLinesP*, que identifica segmentos de reta no espectro binarizado.

A análise dos ângulos obtidos envolve a identificação e remoção de valores discrepantes, seguindo um critério baseado no desvio padrão e utilizando o método estatístico *Z-Score*, de forma semelhante ao método baseado na dilatação dos pixels. O ângulo predominante da inclinação é determinado como a média dos ângulos restantes após a filtragem dos valores extremos.

Por fim, a imagem original é rotacionada no sentido oposto ao ângulo identificado, corrigindo a inclinação e alinhando os caracteres horizontalmente, garantindo uma melhor normalização da orientação do documento.

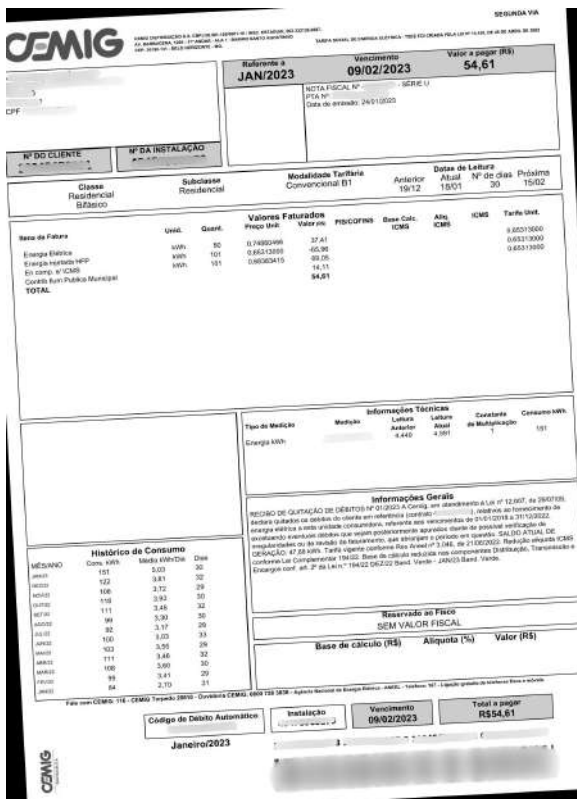
2.3.3 Resultados e Discussão

Para avaliar os algoritmos de detecção de orientação propostos, foram utilizados dois *layouts* de notas fiscais de energia elétrica. A partir dos layouts com orientação conhecida, foram geradas novas imagens com orientações que variam de -30° até 30° , com passo de $0,2^\circ$, totalizando 600 imagens. Alguns exemplos de imagens utilizadas, podem ser visualizadas nas Figuras 15, 16 e 17.

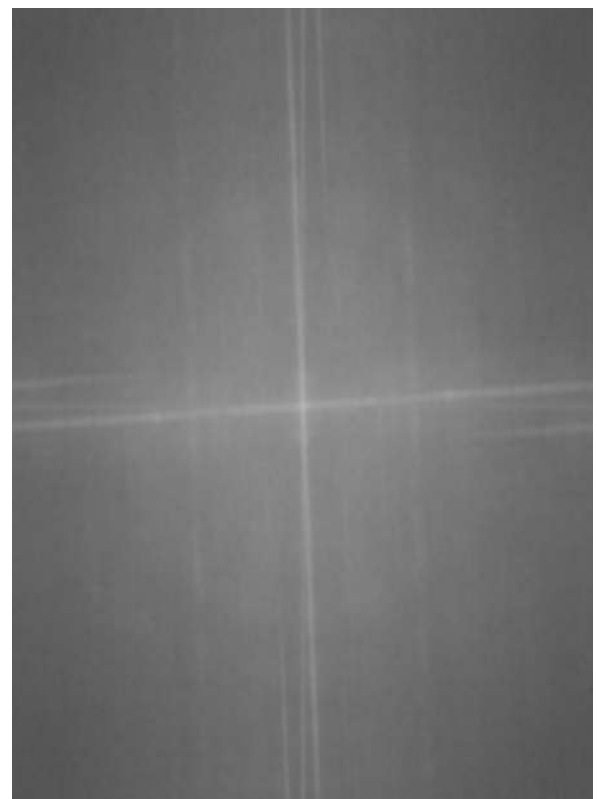
As Tabelas 2 e 3 apresentam os resultados da raiz do erro quadrático médio (RMSE) e do erro médio absoluto (MAE) obtidos na detecção de orientação das imagens do conjunto de dados criado, considerando tanto as faturas individualmente quanto o conjunto completo das imagens analisadas.

Para o método baseado na dilatação dos pixels (Tabela 2), os resultados indicam um RMSE de $0,13938^\circ$ e um MAE de $0,10795^\circ$. O melhor desempenho foi obtido nas variações criadas a partir da imagem 1, com erro médio inferior a $0,1^\circ$. Além disso, esse método apresentou um tempo médio de execução de 0,028 segundos por imagem.

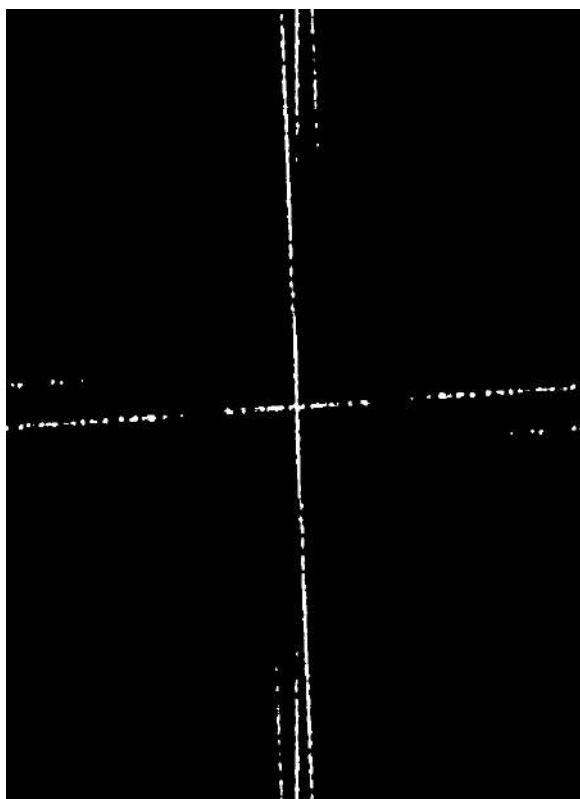
No caso do método baseado no espectro de frequência (Tabela 3), os resultados apontam um RMSE de $0,249^\circ$ e um MAE de $0,210^\circ$. Assim como no primeiro método, o melhor desempenho ocorreu nas variações da imagem 1, com erro absoluto médio de $0,208^\circ$. No entanto, esse método apresentou um tempo médio de detecção de 0,118 segundos por imagem, um valor superior ao do método baseado na dilatação de pixels. Esse aumento no tempo de execução pode ser atribuído à necessidade da conversão de



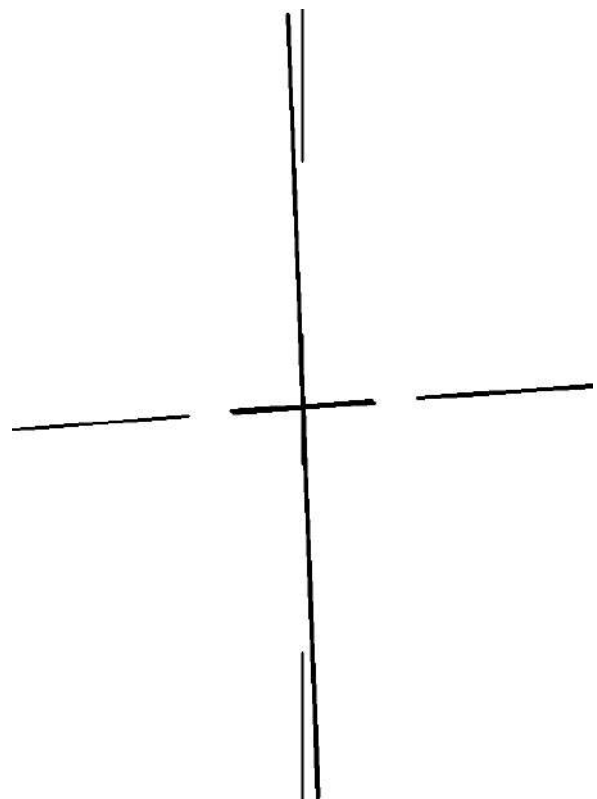
(a) Imagem original no domínio espacial.



(b) Módulo da imagem no domínio da frequência.



(c) Espectro Binarizado.



(d) Linhas detectadas.

Figura 14 – Avaliação da imagem pela análise em frequência.

domínios por meio da *Transformada de Fourier*, além da complexidade da detecção de linhas por meio da *Transformada de Hough*.

Tabela 2 – Resultado detecção de orientação. Método de Dilatação dos Pixels

Imagem	RMSE	MAE
Imagem 1	0,105	0,081
Imagem 2	0,167	0,135
Ambas	0,139	0,108

Tabela 3 – Resultado detecção de orientação. Método de Espectro de Frequência.

Imagem	RMSE	MAE
Imagem 1	0,245	0,208
Imagem 2	0,254	0,212
Ambas	0,250	0,210

Por fim, conclui-se que, embora ambos os algoritmos tenham apresentado baixo erro na detecção da orientação e sejam adequados para essa tarefa, o método baseado na dilatação de pixels demonstrou desempenho superior, com menor erro e tempo de execução, em comparação ao método do espectro da frequência. Dessa forma, o primeiro método apresentado foi selecionado como a abordagem a ser utilizada neste projeto para detecção e correção da orientação. A aplicação dessa técnica ao conjunto de dados desenvolvido pode ser observada nas Figuras 15, 16 e 17, na qual a imagem original é apresentada a esquerda e a imagem corrigida, após a compensação angular, é exibida a direita.

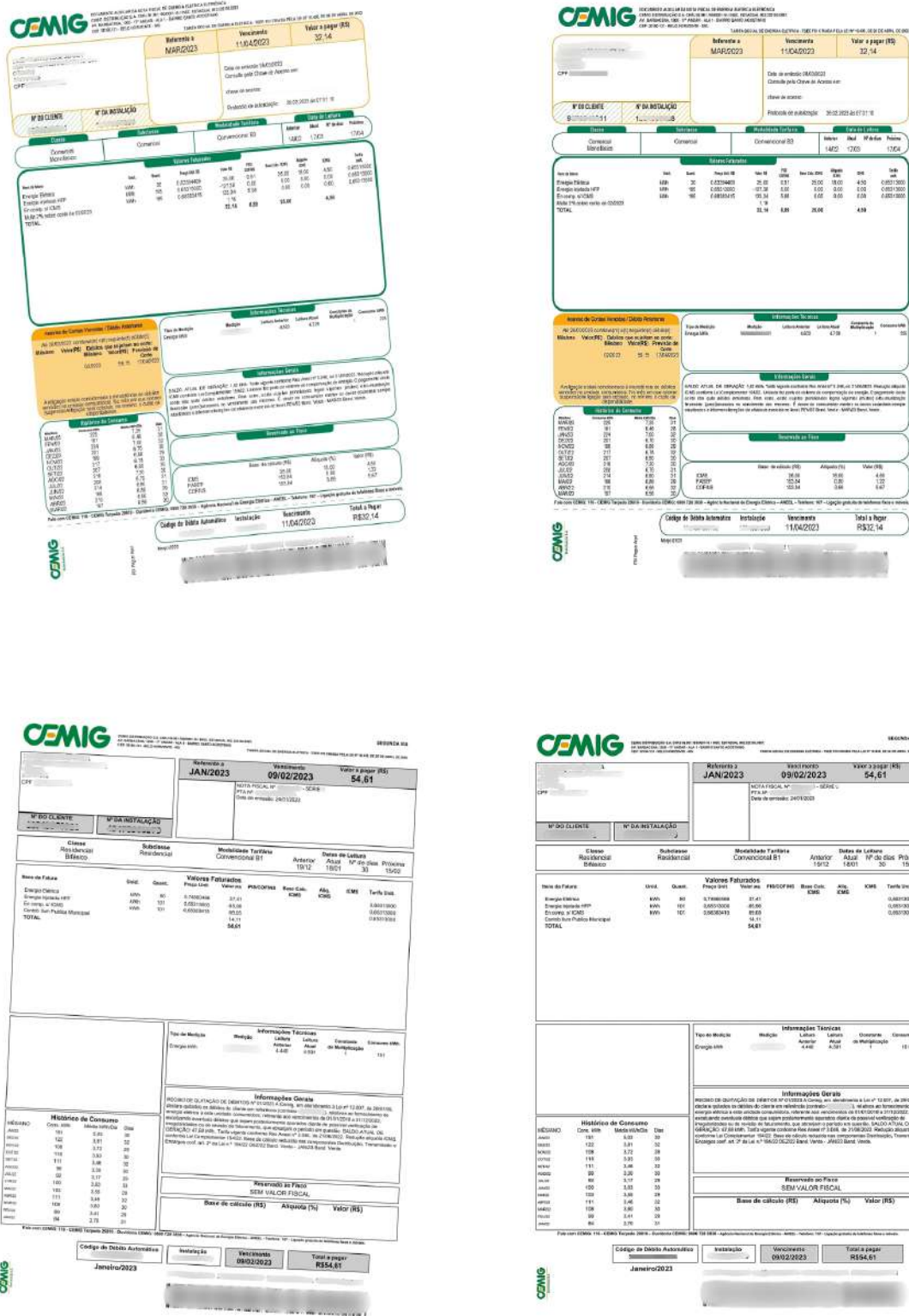


Figura 15 – Aplicação do algoritmo de correção de orientação - Exemplo 1. Imagem original à esquerda e corrigida à direita.

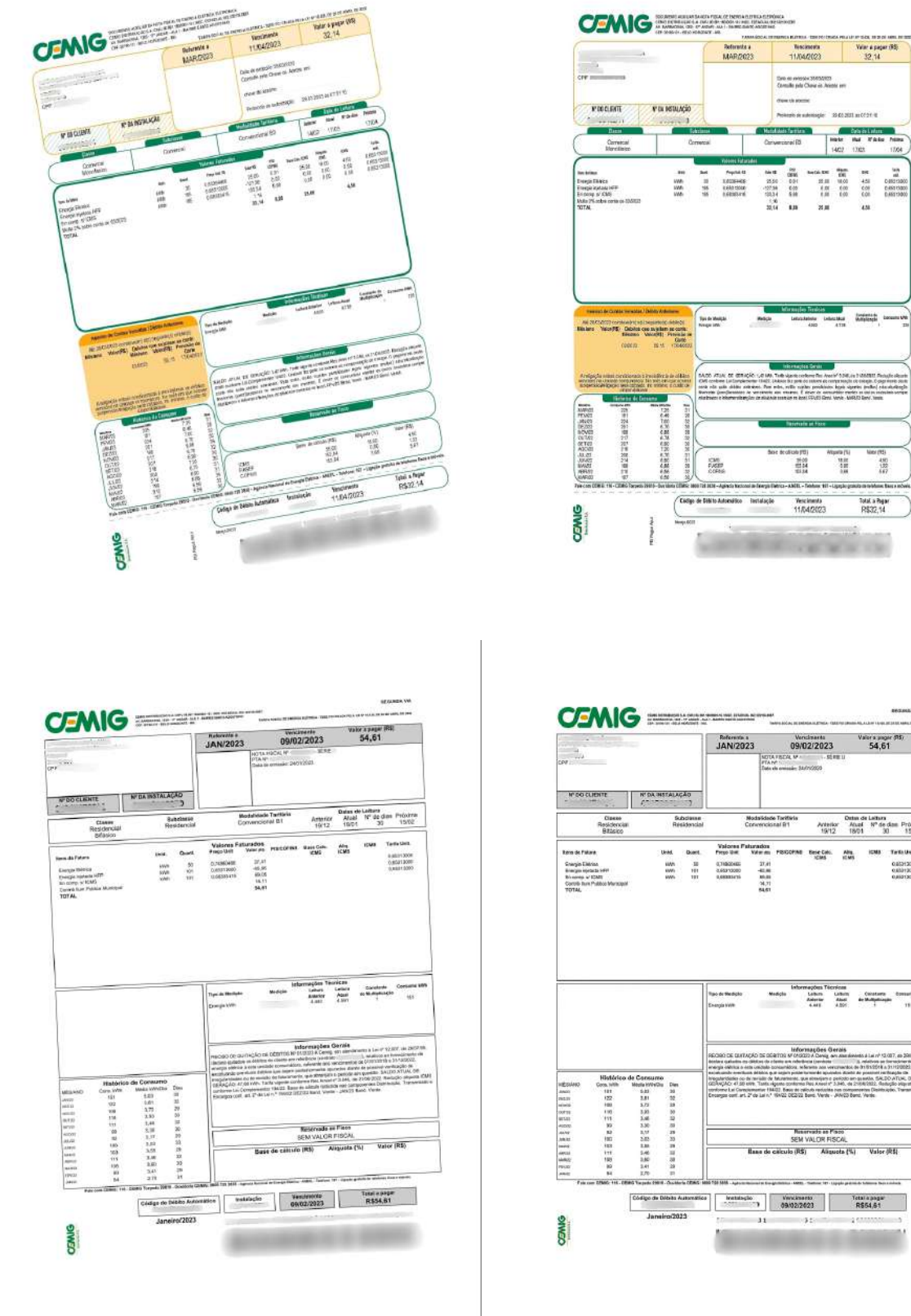


Figura 16 – Aplicação do algoritmo de correção de orientação - Exemplo 2. Imagem original à esquerda e corrigida à direita.

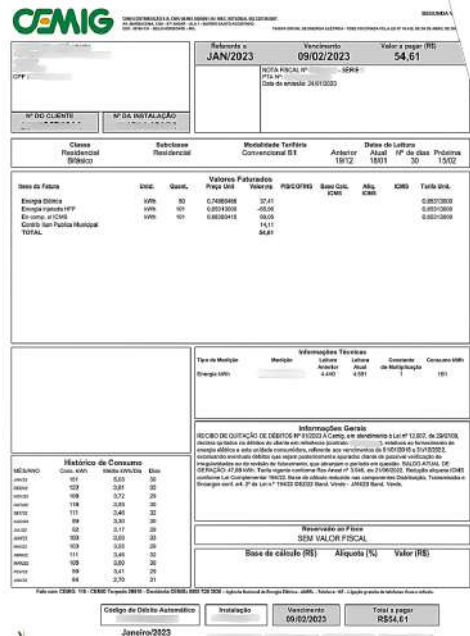
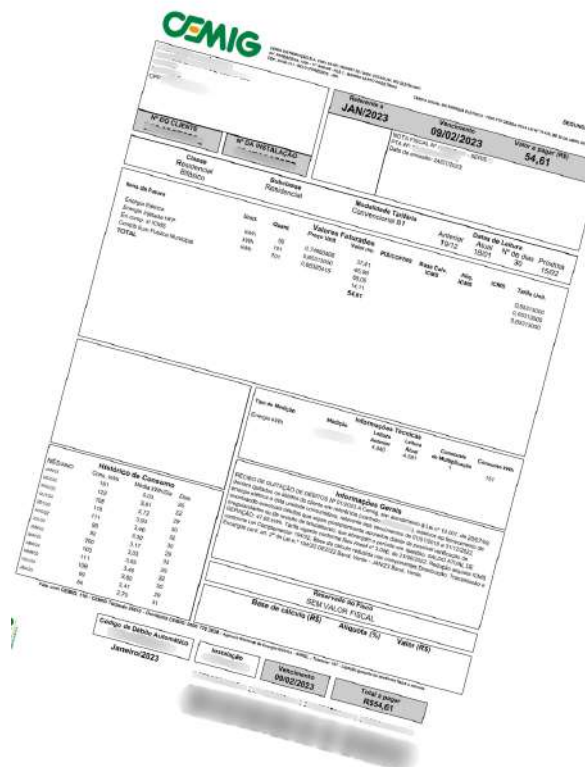
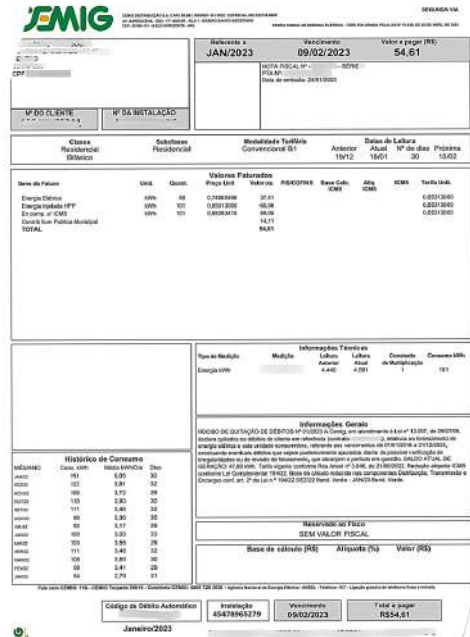
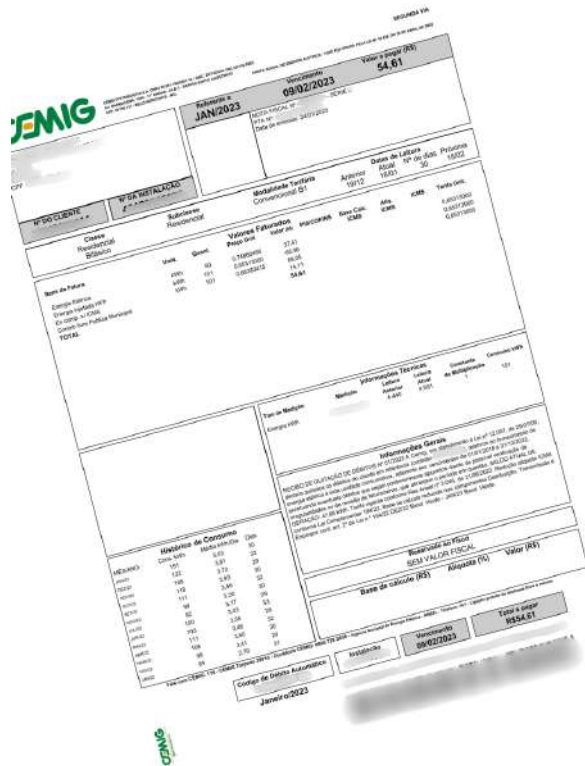


Figura 17 – Aplicação do algoritmo de correção de orientação - Exemplo 3. Imagem original à esquerda e corrigida à direita.

3 Classificação de Concessionárias pela Identificação Automática das Faturas

A classificação e análise automática de faturas representam um avanço significativo na automação de processos organizacionais, permitindo maior agilidade e eficiência na gestão documental. Documentos não estruturados, como faturas, são frequentemente utilizados como registros de transações e comprovação de pagamentos, tornando essencial seu processamento adequado. A realização manual dessas tarefas, além de demandar tempo considerável, pode resultar em erros e atrasos. Nesse contexto, a extração automática de dados emerge como uma solução eficiente, automatizando tarefas repetitivas e proporcionando um aumento expressivo na produtividade (BAVISKAR et al., 2021).

Com a crescente digitalização e automação de processos, a identificação automática de layouts de documentos tornou-se uma necessidade para diversas empresas, especialmente aquelas que lidam com grandes volumes de dados. A manutenção dos documentos em formato digital elimina a necessidade de impressão e organização manual em arquivos físicos, contribuindo para a modernização dos fluxos de trabalho e a redução de custos operacionais (NASIR et al., 2020). Segundo Jiang et al. (2022), grandes empresas de tecnologia enfrentam desafios crescentes no gerenciamento e organização de documentos técnicos produzidos por engenheiros e gestores. Esse cenário evidencia a necessidade de métodos eficientes de classificação automática de documentos, capazes de otimizar processos e garantir a acessibilidade das informações.

Uma abordagem amplamente explorada na literatura para classificação de documentos baseia-se no uso de Redes Neurais Convolucionais (CNNs). Kang et al. (2014) propõem o uso de CNNs para a classificação de imagens de documentos, destacando que a similaridade estrutural entre classes de documentos justifica a escolha dessa arquitetura. Segundo os autores, os experimentos realizados demonstraram a eficácia da abordagem proposta, reforçando a capacidade das CNNs de reconhecer e classificar padrões estruturais presentes em imagens (BHATT et al., 2021).

As CNNs são amplamente reconhecidas como ferramentas eficazes para a extração de recursos visuais em imagens, sendo consideradas uma das arquiteturas de aprendizado profundo mais populares (LIU; PU; SUN, 2021). Além de sua capacidade de extrair características relevantes, as CNNs apresentam alta generalização, tornando-as adequadas para diversas aplicações (LIU; PU; SUN, 2021). No entanto, determinar a arquitetura ideal para uma aplicação específica é um desafio. Tan e Le (2019) demonstram que o equilíbrio entre profundidade, largura e resolução da rede pode levar a um melhor desempenho, tornando-se um fator determinante na eficácia do modelo.

Nesse sentido, [Tan e Le \(2019\)](#) propõem a família de modelos *EfficientNets*, que alcançam maior precisão e eficiência em comparação às arquiteturas anteriores de CNNs. Os autores destacam que o modelo *EfficientNet-B7* atinge uma precisão de 84,4% no top-1 e 97,1% no top-5 do conjunto *ImageNet*, sendo 8,4 vezes menor e 6,1 vezes mais rápido em inferência do que o melhor modelo de *ConvNet* disponível. Esse desempenho ressalta a importância de escolher arquiteturas otimizadas para aplicações de classificação de documentos.

Além da arquitetura do modelo, a qualidade do conjunto de dados utilizado no treinamento e validação é um fator essencial. Segundo [Guérin et al. \(2021\)](#), o *ImageNet* é um dos mais amplos e diversificados conjuntos de dados públicos rotulados, sendo amplamente utilizado como extrator de recursos para diferentes tarefas de visão computacional. Em seu estudo sobre agrupamento de imagens (*Image Clustering - IC*), [Guérin et al. \(2021\)](#) investigam as relações entre o conjunto de dados, a arquitetura CNN, a camada de extração de recursos e o algoritmo de agrupamento. O objetivo do IC é agrupar imagens por similaridade de conteúdo, separando objetos distintos em diferentes categorias. Nesta linha, os resultados apresentados por [Guérin et al. \(2021\)](#) indicam que a última camada antes da camada *softmax* é a mais eficiente para a extração de recursos. Além disso, os autores destacam que a escolha da arquitetura da CNN tem um impacto significativo na qualidade dos agrupamentos, tornando-se um fator crucial para o sucesso da tarefa.

Tarefas como classificação e agrupamento de documentos podem ser aprimoradas por meio de métricas de similaridade bem definidas ([XIA; ZHANG; LI, 2015](#)). A medida de similaridade é uma ferramenta fundamental para determinar o grau de correspondência entre dois objetos ([YE, 2011](#)). Na literatura, diversas abordagens são propostas para essa análise. A distância euclidiana, uma das métricas mais simples e frequentemente utilizada em aprendizado de máquina, apresenta limitações ao gerar representações discriminativas ([XIA; ZHANG; LI, 2015](#)). Por outro lado, a Similaridade de Cosseno, amplamente aplicada em espaços de alta dimensão, foca na orientação dos vetores em vez da magnitude, permitindo uma comparação mais eficiente entre documentos. Essa métrica baseia-se no cálculo do cosseno do ângulo entre dois vetores, assumindo que vetores semelhantes apresentam menor ângulo entre si ([XIA; ZHANG; LI, 2015](#)).

Este capítulo aborda a aplicação de CNNs para a extração de características e a classificação de faturas de energia elétrica, com o objetivo de identificar a distribuidora de energia responsável pela emissão do documento. A abordagem considera o layout específico de cada fatura como critério principal de classificação. Adicionalmente, para algumas distribuidoras, propõe-se um segundo método baseado exclusivamente no conteúdo textual extraído da imagem por meio de uma ferramenta de OCR.

A implementação dessas técnicas possibilita que organizações que lidam com grandes volumes de faturas apliquem os métodos desenvolvidos neste trabalho para classificar

automaticamente documentos diversos. Isso viabiliza a automação do processo de extração de dados de acordo com cada tipo de layout, facilitando a integração de sistemas e promovendo a eficiência operacional, além de reduzir erros humanos na gestão documental.

As principais contribuições deste capítulo estão relacionadas à aplicação de técnicas, com enfoque na classificação automatizada de notas fiscais de energia elétrica. Especificamente, destacam-se três principais abordagens: (i) a extração de características a partir de imagens de faturas utilizando CNNs; (ii) a análise de similaridade de vetores de características, permitindo a identificação de padrões entre documentos distintos; e (iii) o desenvolvimento de algoritmos que possibilitam a classificação automática de faturas sem a necessidade de treinar novos modelos.

Ao explorar a aplicação dessas técnicas no processamento de documentos, este estudo não apenas propõe um método eficiente para automatizar a tarefa de classificação, mas também contribui significativamente para a área de processamento inteligente de documentos. A abordagem apresentada amplia a aplicabilidade das técnicas de aprendizado profundo e análise de similaridade, viabilizando soluções mais escaláveis e adaptáveis a diferentes cenários empresariais e institucionais.

3.1 CNNs na Extração de Características e Classificação de Imagens

A extração de características em imagens por meio de Redes Neurais Convolucionais (CNNs) tem sido amplamente utilizada em diversas áreas, incluindo o diagnóstico médico, autenticação biométrica e classificação de documentos. A seguir, são apresentados estudos que demonstram a aplicabilidade das CNNs nesses contextos, com ênfase na extração de recursos, análise de similaridade e classificação de imagens.

No campo da medicina, a caracterização automatizada de tumores tem sido um dos principais desafios da área de diagnóstico auxiliado por computador. Diante da limitação na classificação de imagens médicas devido ao tamanho reduzido dos conjuntos de dados, [Deepak e Ameer \(2021\)](#) propõem a combinação de CNNs com Máquinas de Vetores de Suporte (SVM) para melhorar a classificação dessas imagens. Os resultados indicam que, quando os dados de treinamento são limitados, a SVM apresenta um desempenho superior ao classificador *softmax* para as *features* extraídos pela CNN.

Ainda no contexto de diagnóstico médico, [Jafari e Karami \(2023\)](#) apresentam um método baseado na extração e redução de características para a detecção de câncer de mama em imagens de mamografia. A abordagem utiliza CNNs pré-treinadas para extrair características que são posteriormente combinadas e reduzidas, sendo classificadas por algoritmos como Redes Neurais (NN), k-Nearest Neighbors (kNN), Random Forest (RF) e SVM. Os resultados demonstram a superioridade do método em comparação com técnicas

convencionais, evidenciando sua eficácia no diagnóstico de lesões mamárias.

Outra aplicação relevante das CNNs no setor da saúde está relacionada à detecção de infartos isquêmicos agudos. [Tasci \(2023\)](#) propõem um método utilizando CNNs pré-treinadas, como EfficientNet B0, DenseNet201, ResNet101, ResNet50, InceptionResNetV2, Xception, MobileNetV2, ShuffleNet, DarkNet19, NASNetLarge e AlexNet. As características extraídas das camadas *softmax* das CNNs foram classificadas com SVM, e os vetores de predição foram combinados utilizando o método *Iterative Majority Voting* (IMV). Segundo os autores, o método permite classificar com precisão infartos isquêmicos agudos em imagens de ressonância magnética de difusão, auxiliando neurologistas na detecção da condição e reduzindo erros de diagnóstico.

Além do diagnóstico médico, CNNs também têm sido empregadas para aprimorar a identificação de patologias gastrointestinais. [Noor et al. \(2023\)](#) propõem um método baseado em um mecanismo de atenção para destacar áreas afetadas em imagens endoscópicas. Imagens sobrepostas são geradas combinando as imagens originais com aquelas realçadas pelo mecanismo de atenção. As características extraídas por CNNs são reduzidas por meio de um algoritmo de seleção baseado em similaridade de cosseno e, posteriormente, classificadas por algoritmos de aprendizado de máquina. Os autores afirmam que essa abordagem pode tornar a endoscopia uma ferramenta mais acessível e eficiente para o diagnóstico de doenças gastrointestinais.

No campo da autenticação biométrica, [Ebrahimpour \(2023\)](#) apresentam um sistema de reconhecimento de íris baseado na arquitetura *MobileNet*. O modelo extrai características das imagens de íris e compara os vetores gerados com aqueles armazenados no banco de dados. A identidade do indivíduo é confirmada caso a distância entre os vetores seja menor que um limiar pré-estabelecido. Os resultados indicam que o método proposto apresenta alta precisão para identificação e autenticação, destacando a eficiência das CNNs na extração de padrões biométricos.

No que se refere à classificação de documentos, que é o foco deste trabalho, [Jiang et al. \(2022\)](#) argumentam que estudos anteriores se concentraram predominantemente no processamento de texto, apesar de documentos técnicos frequentemente conterem informações multimodais. Para lidar com esse desafio, os autores apresentam a arquitetura de aprendizado profundo multimodal *TechDoc*, desenvolvida para explorar características tanto visuais quanto textuais em documentos. Os resultados indicam que o *TechDoc* supera os métodos unimodais de última geração em termos de precisão de classificação.

Outro estudo relevante na área de classificação de documentos é o de [Nasir et al. \(2020\)](#), que apresentam uma técnica de aprendizado supervisionado em tempo real baseada em redes neurais convolucionais profundas (DCNN). A abordagem inclui etapas de aumento de dados, extração de características por CNNs pré-treinadas, fusão e seleção de recursos. Testado no conjunto de dados *Tobacco3482*, o método atingiu uma precisão

de 93,1%, demonstrando sua viabilidade para aplicações em ambientes empresariais.

Por fim, [Harley, Ufkes e Derpanis \(2015\)](#) propõem um novo estado da arte para classificação e recuperação de imagens de documentos utilizando redes neurais convolucionais profundas. Os autores destacam que, na análise de objetos e cenas, as CNNs são capazes de aprender uma hierarquia de abstrações. Experimentos demonstram que as características extraídas pelas CNNs são robustas à compressão, além de indicarem que redes treinadas em imagens não documentais podem ser transferidas para tarefas de análise de documentos sem necessidade de treinamento adicional para aprendizado de características específicas.

Dessa forma, os estudos apresentados demonstram a versatilidade das CNNs na extração de características e na classificação de imagens em diferentes domínios. No contexto do presente trabalho, esses avanços fundamentam a aplicação das CNNs na classificação automatizada de notas fiscais, permitindo a identificação precisa de layouts distintos e viabilizando a integração com métodos de análise de similaridade de vetores de características.

3.2 Classificação de documentos por imagem e por texto

A classificação automática de documentos pode ser realizada por diferentes abordagens, dependendo da natureza das informações contidas nos arquivos. Este trabalho propõe dois métodos complementares para a categorização de faturas de energia elétrica: um baseado na análise da imagem do documento, utilizando redes neurais convolucionais (CNNs), e outro fundamentado na extração e análise do texto, empregando técnicas de *Fuzzy Matching*. A seguir, detalham-se ambos os métodos e suas respectivas aplicações.

3.2.1 O Desenvolvimento da Proposta

A abordagem baseada na análise de imagem utiliza modelos da família *EfficientNet*, uma rede neural convolucional previamente treinada no conjunto de dados *ImageNet*. Para aplicar essa técnica à classificação de faturas, a camada de saída da *EfficientNet* é removida, permitindo que o modelo funcione exclusivamente como um extrator de características visuais.

O primeiro passo do processo consiste no pré-processamento das imagens dos documentos, incluindo redimensionamento e normalização, a fim de garantir compatibilidade com a arquitetura da rede. Em seguida, as características visuais extraídas do documento são representadas por um vetor de características, que será comparado com vetores previamente armazenados, associados a classes conhecidas de faturas.

Para isso, através da biblioteca Keras, o modelo foi carregado com o parâmetro `include_top`, como `False`, o que indica que a camada totalmente conectada no topo da

rede (*fully connected layer*), não foi incluída.

Para otimizar o tempo de inferência, os vetores de referência são extraídos antecipadamente e armazenados em arquivos *.npy*, garantindo que, durante a classificação, apenas o documento de entrada tenha suas características extraídas em tempo real. A escolha dessas imagens utilizadas como referência foi aleatória.

A similaridade entre os vetores é medida por meio da distância de cosseno, uma métrica que avalia o grau de semelhança entre dois vetores. A equação para o cálculo da distância de cosseno entre dois vetores \mathbf{u} e \mathbf{v} é definida como:

$$d_{\cos} = 1 - \frac{\mathbf{u} \cdot \mathbf{v}}{|\mathbf{u}||\mathbf{v}|}. \quad (3.1)$$

A classe atribuída ao documento de entrada corresponde àquela cujo vetor de referência apresenta a menor distância de cosseno em relação ao vetor extraído do documento. Essa metodologia permite uma categorização eficiente sem a necessidade de treinar novos modelos, tornando o processo mais ágil e escalável. O funcionamento do método proposto pode ser visualizado no Algoritmo 2.

Algoritmo 2 Classificação de Documentos por Extração de Características

Entrada: Documento a ser classificado

Parâmetros: Vetores de características armazenados

Saída: Classe do documento

- 1: Redimensionar e normalizar imagem.
 - 2: Extrair o vetor de características da imagem.
 - 3: Iterar sobre vetores de características armazenados
 - 4: Calcular a similaridade de cosseno entre o vetor extraído e o vetor armazenado.
 - 5: Identificar a classe com vetor mais similar.
 - 6: Retornar classe do documento classificado.
-

Embora a análise de imagem seja uma ferramenta essencial para a classificação de documentos, há situações em que a estrutura visual não fornece informações suficientes para determinar a classe do documento. Em particular, as faturas de energia elétrica podem apresentar variações no layout, mas manter padrões textuais característicos da distribuidora emissora. Para esses casos, este trabalho propõe um segundo método de classificação, baseado na extração e análise do conteúdo textual dos documentos.

A diversidade de modalidades e grupos tarifários nas faturas de energia elétrica resulta em um alto volume de informações a serem processadas. Assim, faz-se necessária uma abordagem adaptativa que possa classificar os documentos de forma precisa sem a necessidade de treinar modelos de aprendizado profundo. O método proposto realiza essa classificação por meio de um sistema baseado em Reconhecimento Óptico de Caracteres (OCR) e técnicas de *Fuzzy Matching* (FM), permitindo identificar padrões textuais mesmo quando há ruídos ou variações na extração do texto.

O processo inicia-se com a conversão da imagem em texto utilizando a ferramenta de OCR, como o *Tesseract*. Em seguida, a técnica de *Fuzzy String Matching* é aplicada para comparar o texto extraído com um conjunto de expressões pré-definidas, representando diferentes classes de faturas. Essa abordagem é especialmente útil em cenários onde a extração textual não é perfeitamente precisa, pois o FM permite identificar correspondências aproximadas entre palavras e frases.

A classificação final é realizada atribuindo ao documento a categoria cuja expressão pré-definida apresentar a maior pontuação de similaridade. O método proposto não requer o treinamento de modelos complexos, reduzindo a necessidade de grandes volumes de dados rotulados e simplificando a implementação.

Apesar de sua simplicidade e eficiência, a técnica de *Fuzzy Matching* pode apresentar limitações em comparação com abordagens mais sofisticadas de aprendizado de máquina, especialmente quando há grande variação nos textos extraídos. No entanto, a combinação desse método com a classificação baseada em imagem permite uma categorização mais robusta e flexível, adaptando-se a diferentes layouts e garantindo maior confiabilidade ao sistema de classificação de faturas de energia elétrica.

3.2.2 Resultados e Discussão

A avaliação dos métodos propostos foi realizada utilizando um *dataset* composto por 143 faturas de energia elétrica, distribuídas em 13 diferentes *layouts*. Para a classificação baseada em imagem, 13 dessas faturas foram selecionadas aleatoriamente como referências e, conseqüentemente, excluídas do conjunto de avaliação do algoritmo. Dessa forma, o conjunto final utilizado para testar o desempenho do método foi composto por 130 imagens.

Adicionalmente, para analisar o comportamento do modelo em um cenário com maior variabilidade nas condições das imagens, foram aplicadas técnicas de aumento de dados (*data augmentation*) ao conjunto original. Esse processo incluiu rotações limitadas a um ângulo de ± 6 graus, garantindo que a área externa da imagem fosse suavemente preenchida por replicação dos pixels da borda original, mantendo a coerência visual. Além disso, foram aplicados deslocamentos e escalonamentos de até 5 e 10%, combinados com rotações. Também foram incorporadas modificações na coloração da imagem, incluindo ajustes na matiz, saturação e valor, bem como variações no brilho e no contraste.

Com essa estratégia, para cada imagem do conjunto original foram geradas 10 versões aumentadas, resultando em um *dataset* expandido de 1300 imagens. Esse procedimento possibilitou avaliar a robustez do método frente a variações visuais comuns em documentos digitalizados.

A análise experimental teve como principais métricas a acurácia do algoritmo e

o tempo de classificação das imagens. Além disso, os experimentos foram conduzidos utilizando dois modelos distintos de CNNs, *EfficientNetB0* e *EfficientNetB7*, ambos com pesos pré-treinados no conjunto de dados *ImageNet*, a fim de comparar o impacto do tamanho e da profundidade do modelo na tarefa de classificação.

Algumas das faturas utilizadas como referência para o algoritmo, bem como exemplos das imagens processadas no experimento, podem ser visualizadas nas Figuras 18, 19, 20 e 21. Vale destacar que as imagens originais do banco de dados foram submetidas aos métodos de pré-processamento descritos nos capítulos anteriores deste trabalho, assegurando que a classificação fosse realizada a partir de dados previamente normalizados e ajustados para otimizar a extração de características.

A avaliação do desempenho do método no conjunto de dados original é apresentada na matriz de confusão da Tabela 4. Ao utilizar o modelo *EfficientNetB0*, o algoritmo cometeu apenas um erro, classificando incorretamente uma imagem da classe *cemig3* como *cemig2*. Embora os *layouts* dessas classes não sejam visualmente semelhantes, o algoritmo conseguiu identificar corretamente a distribuidora de energia, diferenciando-a de outras empresas, mas confundindo-se apenas entre os *layouts* específicos da mesma concessionária. O método atingiu uma acurácia de 0,999 no conjunto de dados original e um tempo médio de classificação por fatura de 1,621 segundos.

A Tabela 5 exibe os resultados obtidos quando o método foi aplicado ao conjunto de dados ampliado. Utilizando novamente o modelo *EfficientNetB0*, o algoritmo alcançou uma acurácia de 0,985. A matriz de confusão indica um comportamento semelhante ao do primeiro experimento, com a maioria dos erros de classificação ocorrendo entre diferentes *layouts* da mesma concessionária. Esse padrão reforça a capacidade das redes neurais convolucionais (*CNNs*) de realizar uma análise semântica das imagens, identificando características que permitem associar corretamente as faturas à sua respectiva distribuidora, mesmo diante de variações no *layout*.

Entretanto, nota-se uma leve variação nos resultados de precisão com a introdução de imagens contendo distorções ou condições visuais menos favoráveis. É possível que a diversidade introduzida por essas novas amostras tenha influenciado o desempenho do algoritmo em determinados casos. Os resultados sugerem que o processamento e a correção de distorções podem desempenhar um papel relevante na eficácia do método. No caso do modelo *EfficientNetB7*, observou-se uma redução na acurácia ao se utilizar o conjunto de dados ampliado. Enquanto no conjunto original o modelo atingiu uma acurácia de 0,940, no conjunto com maior diversidade visual o desempenho caiu para 0,885.

As matrizes de confusão apresentadas nas Tabelas 6 e 7 revelam um padrão consistente com os experimentos anteriores. O modelo *EfficientNetB7* conseguiu identificar corretamente a distribuidora *cemig*, mas apresentou dificuldades na diferenciação entre seus diferentes *layouts*. Além disso, observou-se que o uso do modelo *EfficientNetB7*

CEEESC Distribuição S.A. DOCUMENTO AUXILIAR DA NOTA FISCAL DE ENERGIA ELÉTRICA

SEGUNDA VIA

Referência a: **10/05/2024** Vencimento: **18/09/2024** Valor a pagar (R\$): **RS 106,00**

Comunicado importante: Estamos implementando nosso novo sistema comercial para oferecer a você um serviço mais moderno e acessível. Mas fique tranquilo: você não será prejudicado! Qualquer dúvida sobre problemas nas leituras envie-nos no whats no link no topo do boleto e o atendimento da energia costará, sem cobrança de juros e multa, por 90 dias. Qualquer inconveniência será corrigida no mês seguinte.

Letura Anterior	Letura Atual	Dias	Origem da Letura	Problema Letura
10/05/2024	18/09/2024	31	Lida	

Nome de Fatura	Unid.	Quant.	Preço Unit.	Valor	ICMS	Base Calc.	Alíq. (%)	ICMS (%)	Tarifa
Consumo TE	KWH	100,00	0,847520	84,75	1,27	86,48	17,00	0,28	0,000000
Consumo TUSD	KWH	190,00	0,393300	74,73	1,30	83,43	17,00	0,28	0,000000
Energia SCEE via BENTK	MWh	97	0,197131	19,11					0,000000
Energia comprada GDI	MWh	97	0,467330	45,30					0,000000
TOTAL				106,00					

(a) celesc

CEMIG DOCUMENTO AUXILIAR DA NOTA FISCAL DE ENERGIA ELÉTRICA

SEGUNDA VIA

Referência a: **02/10/2023** Vencimento: **26/10/2023** Valor a pagar (R\$): **4.236,64**

Comunicado importante: Estamos implementando nosso novo sistema comercial para oferecer a você um serviço mais moderno e acessível. Mas fique tranquilo: você não será prejudicado! Qualquer dúvida sobre problemas nas leituras envie-nos no whats no link no topo do boleto e o atendimento da energia costará, sem cobrança de juros e multa, por 90 dias. Qualquer inconveniência será corrigida no mês seguinte.

Nome de Fatura	Unid.	Quant.	Preço Unit.	Valor	ICMS	Base Calc.	Alíq. (%)	ICMS (%)	Tarifa
Consumo TE	KWH	100,00	0,847520	84,75	1,27	86,48	17,00	0,28	0,000000
Consumo TUSD	KWH	190,00	0,393300	74,73	1,30	83,43	17,00	0,28	0,000000
Energia SCEE via BENTK	MWh	97	0,197131	19,11					0,000000
Energia comprada GDI	MWh	97	0,467330	45,30					0,000000
TOTAL				4.236,64					

(b) cemig 1

CEMIG DOCUMENTO AUXILIAR DA NOTA FISCAL DE ENERGIA ELÉTRICA

SEGUNDA VIA

Referência a: **02/10/2023** Vencimento: **17/11/2023** Valor a pagar (R\$): **97,14**

Comunicado importante: Estamos implementando nosso novo sistema comercial para oferecer a você um serviço mais moderno e acessível. Mas fique tranquilo: você não será prejudicado! Qualquer dúvida sobre problemas nas leituras envie-nos no whats no link no topo do boleto e o atendimento da energia costará, sem cobrança de juros e multa, por 90 dias. Qualquer inconveniência será corrigida no mês seguinte.

Nome de Fatura	Unid.	Quant.	Preço Unit.	Valor	ICMS	Base Calc.	Alíq. (%)	ICMS (%)	Tarifa
Energia Básica	MWh	82	0,0254303	2,08					0,7400000
Energia SCEE via BENTK	MWh	97	0,197131	19,11					0,0000000
Energia comprada GDI	MWh	97	0,467330	45,30					0,0000000
TOTAL				97,14					

(c) cemig 2

CEMIG DOCUMENTO AUXILIAR DA NOTA FISCAL DE ENERGIA ELÉTRICA

SEGUNDA VIA

Referência a: **02/10/2023** Vencimento: **17/11/2023** Valor a pagar (R\$): **200,53**

Comunicado importante: Estamos implementando nosso novo sistema comercial para oferecer a você um serviço mais moderno e acessível. Mas fique tranquilo: você não será prejudicado! Qualquer dúvida sobre problemas nas leituras envie-nos no whats no link no topo do boleto e o atendimento da energia costará, sem cobrança de juros e multa, por 90 dias. Qualquer inconveniência será corrigida no mês seguinte.

Nome de Fatura	Unid.	Quant.	Preço Unit.	Valor	ICMS	Base Calc.	Alíq. (%)	ICMS (%)	Tarifa
Energia Básica	MWh	100	0,0254303	2,54					0,7400000
Energia SCEE via BENTK	MWh	97	0,197131	19,11					0,0000000
Energia comprada GDI	MWh	97	0,467330	45,30					0,0000000
TOTAL				200,53					

(d) cemig 3

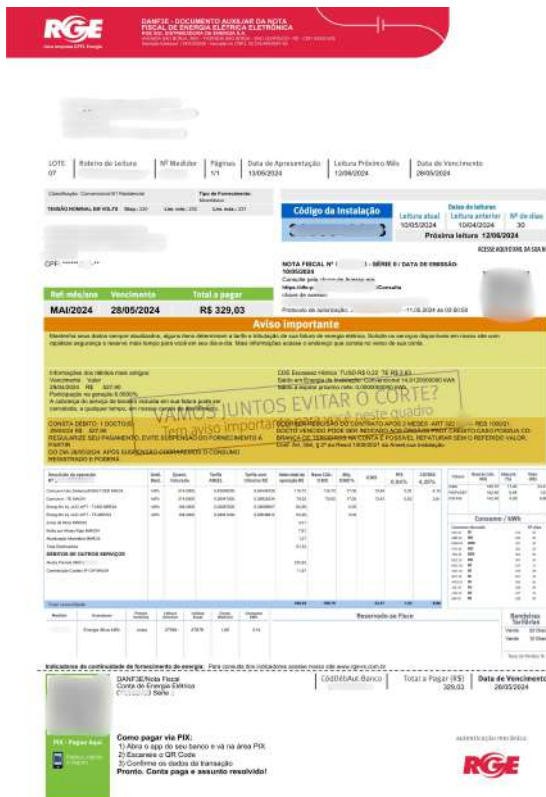
Figura 18 – Imagens usadas para as features de referência - Parte 1.



(a) copel



(b) cpfl



(c) cpfl rge

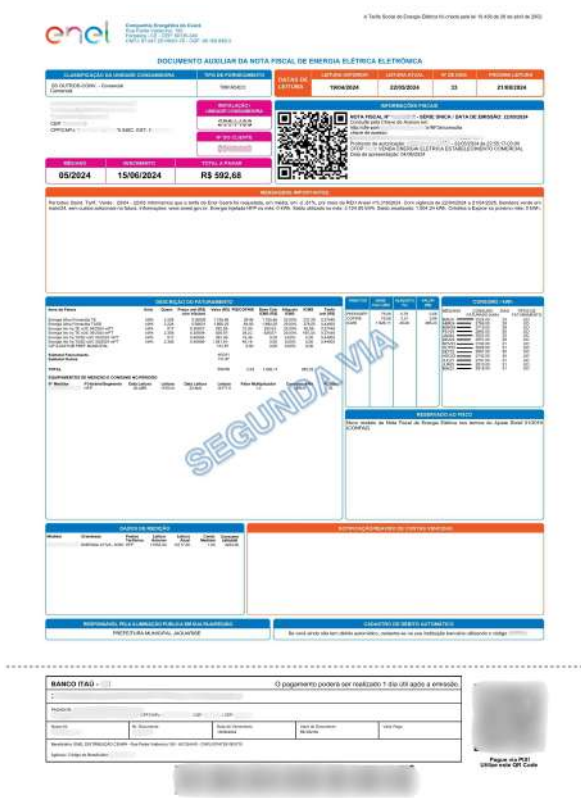


(d) dme

Figura 19 – Imagens usadas para as features de referência - Parte 2.



(a) elektro



(b) enel ce



(c) energisa



(d) equatorial go

Figura 20 – Imagens usadas para as features de referência - Parte 3.



(a) neoenergia pe

Figura 21 – Imagem usada para as features de referência - Parte 4.

resultou não apenas em uma queda na acurácia em relação ao modelo *EfficientNetB0*, mas também em um aumento significativo no tempo médio de classificação por imagem, que passou para 7,15 segundos.

Esses resultados evidenciam o impacto do aumento da complexidade do modelo no tempo de processamento, sem necessariamente garantir uma melhoria proporcional na precisão. A escolha de um modelo mais profundo, como o *EfficientNetB7*, pode trazer bons resultados, mas também demanda maior poder computacional e pode não ser a opção mais eficiente para cenários em que a velocidade de inferência é um fator crítico. Dessa forma, o modelo *EfficientNetB0* demonstrou-se uma alternativa mais equilibrada, oferecendo uma combinação satisfatória entre tempo de processamento e precisão na classificação das faturas de energia elétrica.

O método de classificação baseado em texto foi avaliado utilizando o conjunto de 130 faturas, com o objetivo de identificar a distribuidora de energia responsável por cada documento. No sistema de leitura proposto, essa abordagem é utilizada exclusivamente para a identificação de faturas pertencentes a uma subclasse específica da distribuidora *Cemig*, cuja classificação por imagem não atinge um nível satisfatório de precisão. No entanto, para fins de comparação, o experimento foi conduzido aplicando esse método no mesmo contexto do experimento anterior, permitindo uma análise comparativa entre as abordagens.

Tabela 4 – Matriz de confusão referente ao experimento 1, mostrando o desempenho da classificação de imagens com o modelo *EfficientNetB0*.

	celesc	cemig1	cemig2	cemig3	copel	cpfl	cpfl rge	dme	elektro	enel ce	energisa	equatorial go	neoenergia pe
celesc	10	0	0	0	0	0	0	0	0	0	0	0	0
cemig1	0	10	0	0	0	0	0	0	0	0	0	0	0
cemig2	0	0	10	0	0	0	0	0	0	0	0	0	0
cemig3	0	0	1	9	0	0	0	0	0	0	0	0	0
copel	0	0	0	0	10	0	0	0	0	0	0	0	0
cpfl	0	0	0	0	0	10	0	0	0	0	0	0	0
cpfl rge	0	0	0	0	0	0	10	0	0	0	0	0	0
dme	0	0	0	0	0	0	0	10	0	0	0	0	0
elektro	0	0	0	0	0	0	0	0	10	0	0	0	0
enel ce	0	0	0	0	0	0	0	0	0	10	0	0	0
energisa	0	0	0	0	0	0	0	0	0	0	10	0	0
equatorial go	0	0	0	0	0	0	0	0	0	0	0	10	0
neoenergia pe	0	0	0	0	0	0	0	0	0	0	0	0	10

Diferentemente do experimento realizado com a classificação por imagem, os diferentes *layouts* da *Cemig* foram agrupados, pois não podem ser diferenciados apenas pela análise textual. Esse fator destaca uma limitação do método baseado em texto, evidenciando a vantagem do classificador por imagem na diferenciação de documentos visualmente distintos dentro de uma mesma distribuidora.

Para a identificação das distribuidoras, a análise de similaridade textual foi conduzida utilizando um conjunto de expressões previamente definidas, incluindo os termos: *Celesc*, *Cemig*, *Copel*, *CPFL Paulista*, *RGE Sul*, *DME*, *Elektro*, *Enel*, *Energisa*, *Equatorial* e *Neoenergia Pernambuco*. Essas expressões foram selecionadas com base em sua presença característica nos documentos, permitindo uma correspondência aproximada por meio de técnicas de *Fuzzy Matching*.

Os resultados obtidos são apresentados na Tabela 8. O método baseado em texto alcançou uma acurácia de 92,30%, com um tempo médio de processamento de 2,97 segundos por fatura. Apesar de apresentar um desempenho satisfatório, sua acurácia foi inferior à do método baseado em imagem, especialmente quando comparado ao modelo *EfficientNetB0*, que demonstrou maior robustez na diferenciação dos *layouts*.

Além disso, identificou-se que o método baseado em texto apresenta limitações em cenários específicos. A ausência do cabeçalho em faturas cortadas, por exemplo, compromete a identificação da distribuidora, uma vez que as informações essenciais para a classificação podem não estar visíveis no texto extraído pelo OCR. Da mesma forma,

Tabela 5 – Matriz de confusão referente ao experimento 2, mostrando o desempenho da classificação de imagens com o modelo *EfficientNetB0*.

	celesc	cemig1	cemig2	cemig3	copel	cpfl	cpfl rge	dme	elektro	enel ce	energisa	equatorial go	neoenergia pe
celesc	100	0	0	0	0	0	0	0	0	0	0	0	0
cemig1	0	99	1	0	0	0	0	0	0	0	0	0	0
cemig2	0	0	100	0	0	0	0	0	0	0	0	0	0
cemig3	0	0	15	84	0	0	0	0	0	0	1	0	0
copel	0	0	0	0	100	0	0	0	0	0	0	0	0
cpfl	0	0	0	0	0	100	0	0	0	0	0	0	0
cpfl rge	0	0	0	0	0	0	100	0	0	0	0	0	0
dme	0	0	0	0	0	0	0	98	0	0	2	0	0
elektro	0	0	0	0	0	0	0	0	100	0	0	0	0
enel ce	0	0	0	0	0	0	0	0	0	100	0	0	0
energisa	0	0	0	0	0	0	0	0	0	0	100	0	0
equatorial go	0	0	0	0	0	0	0	0	0	0	0	100	0
neoenergia pe	0	0	0	0	0	0	0	0	0	0	0	0	100

faturas de baixa qualidade ou com artefatos visuais interferem negativamente na extração textual, resultando em falhas na classificação. Esses fatores evidenciam a dependência da integridade do documento para o sucesso da abordagem baseada em texto.

Dessa forma, embora o método textual apresente uma solução viável para casos específicos, sua aplicação deve ser complementada pela abordagem baseada em imagem, garantindo maior precisão e confiabilidade na classificação automática de faturas de energia elétrica. A combinação dessas duas técnicas permite um sistema mais robusto e adaptável a diferentes condições de captura e qualidade dos documentos.

3.2.3 Consideração Finais sobre o Processo de Classificação

Nesta seção, foram exploradas duas abordagens distintas para a classificação de faturas de energia elétrica: um método baseado exclusivamente na análise de imagem e outro fundamentado na extração e processamento do texto presente no documento. Foram conduzidas avaliações experimentais sob diferentes condições para validar a eficácia e a aplicabilidade de cada abordagem.

Uma das principais vantagens dos métodos propostos é a ausência da necessidade de treinamento de novos modelos, o que torna sua implementação mais acessível e eficiente. No método de classificação por imagem, um modelo pré-treinado no conjunto de dados *ImageNet* foi utilizado para extrair características das imagens, enquanto a classificação foi realizada por meio da análise de similaridade entre vetores de características. Essa

Tabela 6 – Matriz de confusão referente ao experimento 3, mostrando o desempenho da classificação de imagens com o modelo *EfficientNetB7*.

	celesc	cemig1	cemig2	cemig3	copel	cpfl	cpfl rge	dme	elektro	enel ce	energisa	equatorial go	neoenergia pe
celesc	10	0	0	0	0	0	0	0	0	0	0	0	0
cemig1	0	7	3	0	0	0	0	0	0	0	0	0	0
cemig2	0	0	10	0	0	0	0	0	0	0	0	0	0
cemig3	0	0	0	10	0	0	0	0	0	0	0	0	0
copel	0	0	0	0	10	0	0	0	0	0	0	0	0
cpfl	1	0	0	0	0	9	0	0	0	0	0	0	0
cpfl rge	0	0	0	0	0	0	10	0	0	0	0	0	0
dme	0	0	0	0	0	0	0	10	0	0	0	0	0
elektro	0	0	0	0	0	0	0	0	8	0	0	2	0
enel ce	0	0	0	0	0	0	0	0	0	10	0	0	0
energisa	0	0	0	0	0	0	0	0	0	0	10	0	0
equatorial go	0	0	0	0	0	0	0	0	0	0	0	10	0
neoenergia pe	0	0	0	0	0	0	0	0	0	0	0	0	10

estratégia reduz significativamente os custos computacionais e o tempo necessário para a implementação do sistema, tornando-o uma solução prática e flexível.

Os resultados obtidos demonstraram a eficiência do uso de CNNs na tarefa de classificação, especialmente com o modelo *EfficientNetB0*, que obteve alto desempenho no conjunto de dados proposto, com precisão de 0,999 e um tempo médio de classificação por fatura de 1,621 segundos. Além disso, os experimentos evidenciaram a capacidade das CNNs de realizar análises semânticas, identificando corretamente a distribuidora de energia na maioria dos casos, mesmo diante de variações no *layout* dos documentos.

Entretanto, verificou-se que distorções nas imagens impactam negativamente a precisão do método, ressaltando a importância do pré-processamento e da correção de artefatos visuais para garantir a confiabilidade do sistema. Além disso, observou-se que, apesar de sua maior complexidade, o modelo *EfficientNetB7* apresentou desempenho inferior ao *EfficientNetB0*, tanto em precisão quanto em tempo médio de classificação por imagem.

O método de classificação baseado em texto, embora mais simples, desempenha um papel essencial quando integrado à abordagem baseada em imagem. Sua aplicação complementa a análise visual, possibilitando um sistema híbrido que melhora a confiabilidade da classificação. A principal vantagem dessa abordagem reside na sua flexibilidade, permitindo a categorização de faturas em cenários onde a informação textual é um critério determinante.

Tabela 7 – Matriz de confusão referente ao experimento 4, mostrando o desempenho da classificação de imagens com o modelo *EfficientNetB7*.

	celesc	cemig1	cemig2	cemig3	copel	cpfl	cpfl rge	dme	elektro	enel ce	energisa	equatorial go	neoenergia pe
celesc	87	0	0	0	11	0	0	0	0	0	2	0	0
cemig1	0	67	28	5	0	0	0	0	0	0	0	0	0
cemig2	0	0	99	1	0	0	0	0	0	0	0	0	0
cemig3	0	1	6	92	0	0	0	0	0	0	0	0	1
copel	1	0	0	0	98	0	0	1	0	0	0	0	0
cpfl	12	0	0	1	3	68	3	0	0	1	2	3	7
cpfl rge	4	0	2	0	1	0	90	2	0	0	0	1	0
dme	6	0	1	0	5	0	0	79	0	3	2	1	3
elektro	0	0	1	0	1	0	1	0	84	0	0	13	0
enel ce	1	0	0	0	2	0	0	2	0	91	0	0	4
energisa	0	0	0	0	1	0	0	0	0	0	99	0	0
equatorial go	2	0	0	0	1	0	0	0	0	0	0	96	1
neoenergia pe	0	0	0	0	0	0	0	0	0	0	0	0	100

Além disso, a combinação das duas técnicas permite otimizar o desempenho do sistema em diferentes condições. Em casos em que a extração de características visuais não é suficiente para uma classificação precisa, a análise textual pode atuar como um critério adicional. Da mesma forma, quando o conteúdo textual extraído não fornece informações suficientes para a categorização do documento, o método baseado em imagem assume um papel mais relevante, analisando o *layout* da fatura.

Dessa forma, a integração dos dois métodos resulta em um sistema mais robusto e adaptável, capaz de lidar com os desafios encontrados no processamento automático de faturas de energia elétrica. Essa abordagem híbrida se apresenta como uma solução eficaz para a detecção e categorização desses documentos, ampliando a confiabilidade e a aplicabilidade do sistema em diferentes contextos.

3.3 Qualidade da Imagem e o Impacto no Pré-processamento

A avaliação da qualidade das imagens é essencial para garantir a precisão dos processos de extração automatizada de informações. Imagens de baixa qualidade podem comprometer a eficiência dos modelos de Reconhecimento Óptico de Caracteres (OCR), dificultando a leitura e interpretação dos textos extraídos. Neste trabalho, após as etapas de processamento da imagem, propõe-se uma metodologia para avaliar sua qualidade, permitindo ajustes no processamento ou, quando necessário, bloqueando a extração de informações para evitar erros.

Tabela 8 – Matriz de confusão referente ao experimento 5, mostrando o desempenho da classificação por texto.

	celesc	cemig	copel	cpfl	cpfl rge	dme	elektro	enel ce	energisa	equatorial go	neoenergia pe
celesc	10	0	0	0	0	0	0	0	0	0	0
cemig	0	25	0	1	0	0	0	0	0	0	4
copel	0	0	10	0	0	0	0	0	0	0	0
cpfl	0	0	0	10	0	0	0	0	0	0	0
cpfl rge	0	0	0	1	9	0	0	0	0	0	0
dme	0	0	0	0	0	9	0	0	0	0	1
elektro	0	0	0	0	0	0	10	0	0	0	0
enel ce	0	0	0	0	0	0	0	9	0	0	1
energisa	0	0	0	0	0	0	0	0	9	0	1
equatorial go	0	0	0	0	0	0	0	0	0	10	0
neoenergia pe	0	0	0	0	0	0	0	0	0	0	10

3.3.1 Metodologia para Avaliação da Qualidade da Imagem

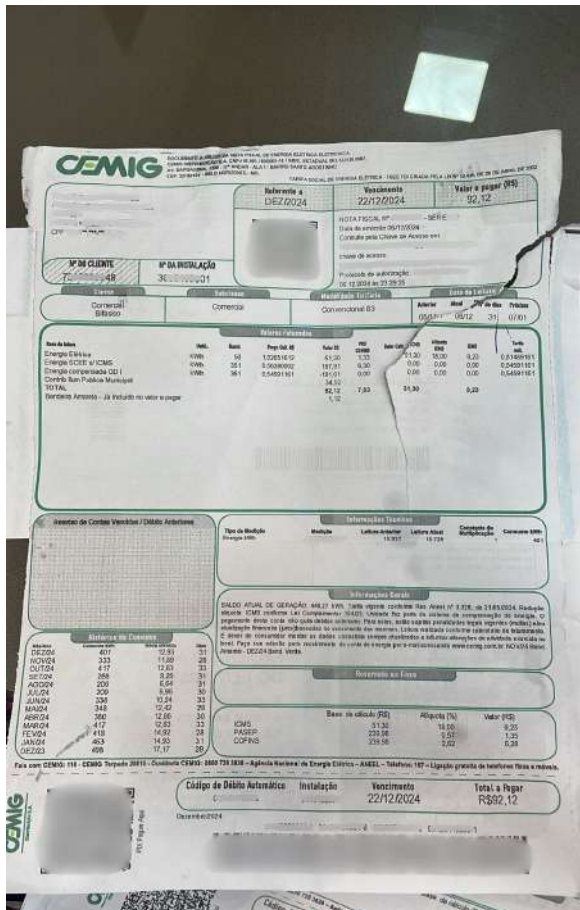
O método proposto baseia-se na análise da confiança das caixas de texto geradas por modelos OCR. Esses modelos, treinados para identificar textos em diversas condições, atribuem um nível de confiança a cada trecho extraído, fornecendo uma estimativa da precisão da leitura realizada. Dessa forma, ao analisar os valores de confiança, é possível correlacionar a qualidade da imagem com a precisão da extração textual.

Após a aplicação das técnicas de pré-processamento, o sistema de OCR (*Tesseract*) é utilizado para extrair o texto da imagem e gerar os respectivos níveis de confiança. Com essas informações, realiza-se uma análise comparativa entre a imagem original e a imagem processada. Inicialmente, calcula-se a média dos valores de confiança de cada versão da imagem, permitindo representar numericamente sua qualidade. Em seguida, são aplicados dois limiares pré-definidos. O limiar mínimo determina a qualidade mínima aceitável para que a extração de informações possa ser realizada, enquanto o limiar médio diferencia imagens de qualidade moderada e alta, auxiliando na tomada de decisão.

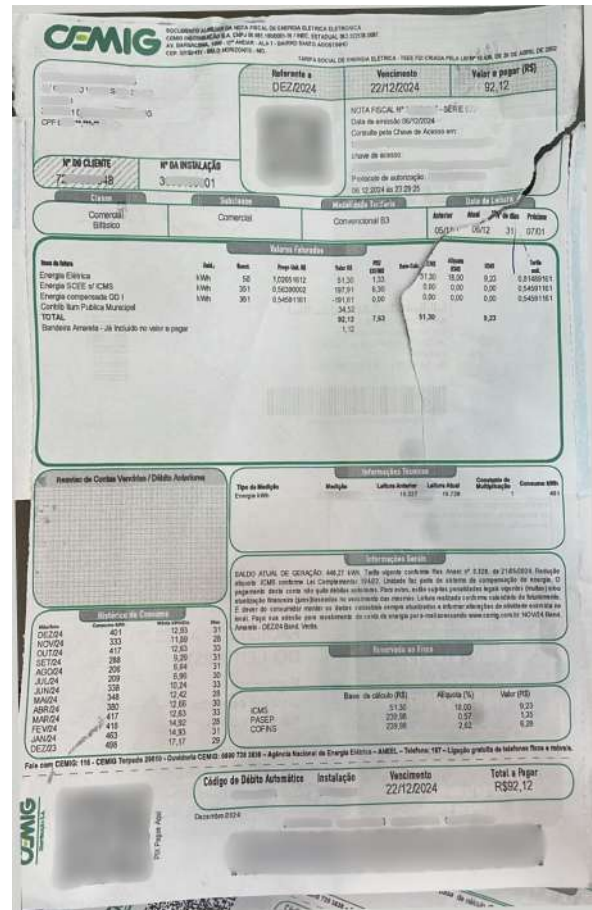
Caso ambas as versões da imagem sejam classificadas como de qualidade inadequada, o sistema pode interromper o processo de extração de informações, assegurando a integridade dos dados. Caso contrário, a versão da imagem com maior média de confiança é selecionada para a etapa subsequente. Esse mecanismo garante que, mesmo quando o processamento da imagem não melhora sua qualidade, a melhor versão disponível seja utilizada.

Neste estudo, considera-se que uma imagem ideal deve apresentar uma confiança

média superior a 40%. A Figura 22 ilustra um exemplo de imagem com variação na orientação e perspectiva, na qual a confiança média da extração de texto foi de 34,45%. Após a aplicação do método de correção, houve um aumento na confiança de 8,83 pontos percentuais, resultando em uma nova imagem com confiança média de 43,28%.



(a) Imagem Original



(b) Imagem Corrigida

Figura 22 – Comparação de nível de confiança entre imagem original e corrigida.

Dessa forma, em um sistema automatizado, a comparação entre as imagens permite identificar cenários em que o processamento não foi eficaz, ao mesmo tempo em que possibilita a continuidade da extração de informações com a versão que apresentar melhor qualidade.

3.3.2 Resultados e Discussão

Para avaliar o impacto do método proposto, foi desenvolvido um *dataset* contendo 115 imagens, representando diferentes cenários de qualidade. O conjunto de dados inclui tanto imagens distorcidas, que necessitam de pré-processamento, quanto imagens em formato ideal, sem necessidade de correções. Todas as imagens foram submetidas às técnicas de pré-processamento descritas nos capítulos anteriores.

O objetivo do experimento foi verificar se os métodos de processamento de imagens são capazes de melhorar a qualidade das imagens, utilizando três métricas principais: a taxa de melhora, o aumento médio de confiança e a diminuição média de confiança. A taxa de melhora corresponde ao percentual de imagens em que o pré-processamento resultou em um aumento na confiança da leitura realizada pelo OCR. O aumento médio de confiança representa o ganho percentual médio na confiança das imagens que apresentaram melhoria. Já a diminuição média de confiança refere-se à magnitude percentual média da redução de qualidade nas imagens que sofreram impacto negativo.

Os resultados indicaram que o algoritmo obteve uma taxa de melhora de 79,13%, ou seja, em 91 das 115 imagens avaliadas, houve um aumento na confiança média do OCR. Entre essas imagens, o aumento médio foi de 5,18% na confiança. Por outro lado, em 24 imagens (20,87% do total), o pré-processamento impactou negativamente a confiança média, resultando em uma diminuição de 1,16%. Esse impacto, entretanto, foi considerado pouco significativo. No geral, ao considerar todas as imagens do conjunto, o algoritmo resultou em um aumento médio de 3,85% na confiança do OCR.

A análise também revelou que as imagens que sofreram redução de qualidade já apresentavam uma confiança inicial elevada, com média de 63,09%, sugerindo que o pré-processamento teve impacto limitado nesses casos. Em contrapartida, para as imagens que tiveram uma melhora, a confiança inicial era significativamente menor, com média de 48,20%, o que reforça a efetividade do método em cenários de baixa qualidade.

Além disso, observou-se que a melhor imagem sem pré-processamento atingiu uma confiança média de 78,25%, enquanto a pior imagem apresentou apenas 24,23% de confiança. A maior taxa de aumento de qualidade registrada foi de 22,52%, enquanto a maior taxa de diminuição foi de 4,04%.

Esses resultados demonstram que o método proposto é eficaz na melhoria de imagens com baixa qualidade, garantindo que documentos degradados possam ser corrigidos antes da extração de informações. Além disso, o sistema de avaliação automática permite selecionar a melhor versão da imagem disponível, seja ela original ou corrigida, assegurando que apenas imagens com qualidade satisfatória sejam utilizadas na análise posterior.

4 Extração de Informações dos Documentos

A extração automatizada de informações de documentos tem sido amplamente explorada por meio de técnicas avançadas de Processamento de Linguagem Natural (NLP) e Reconhecimento Óptico de Caracteres (OCR). Modelos de NLP são eficazes na compreensão textual, mas frequentemente negligenciam a estrutura visual e o layout dos documentos, aspectos fundamentais para uma interpretação precisa das informações contidas nesses arquivos. Para superar essa limitação, [Xu et al. \(2020\)](#) apresentam o *LayoutLM*, uma abordagem que modela a interação entre texto e layout, permitindo uma extração mais contextualizada dos dados.

[Bajrami et al. \(2023\)](#) comparam o desempenho dos modelos *LayoutLM* e *Donut*, ambos baseados em *transformers*, na extração de informações de documentos. Os resultados indicam que o *LayoutLM* obteve uma precisão de 0.88 no conjunto de dados utilizado, enquanto o *Donut* apresentou um desempenho inferior, com precisão de 0.74. O modelo *Donut*, proposto por [Kim et al. \(2021\)](#), adota uma abordagem sem OCR e utiliza um *Transformer* puro para processar imagens de documentos. Essa arquitetura reduz a complexidade do processo ao eliminar a dependência de OCR, mantendo a precisão.

No contexto de extração de informações visuais, o *Visual Question Answering (VQA)* é uma abordagem de aprendizado profundo multimodal que permite responder a perguntas baseadas em texto sobre uma imagem. Existem diferentes tarefas de VQA definidas conforme diversos cenários de aplicação, incluindo gráficos e imagens do cotidiano. Neste contexto, o *Document Visual Question Answering (DVQA)* tem se destacado como uma abordagem multimodal para responder a perguntas sobre documentos ([DING et al., 2022](#)). A pesquisa de [Mishra et al. \(2019\)](#) introduz o *OCR-VQA*, um desafio que combina reconhecimento óptico de caracteres e VQA, possibilitando a leitura de texto diretamente de imagens para responder a questões específicas. Apesar dos avanços, desafios permanecem, especialmente em documentos com múltiplas páginas, como destacado por [Kang et al. \(2024\)](#). Os autores propõem um modelo baseado no *Pix2Struct* para processar documentos extensos de forma mais eficiente, reduzindo a necessidade de grande capacidade computacional e mantendo a precisão da extração.

Diante dessas considerações, este trabalho adota o serviço de consultas personalizadas do *Amazon Textract* para a tarefa de DVQA, realizando o ajuste fino (*fine-tuning*) do modelo com imagens de notas fiscais de energia elétrica. Após testes preliminares comparando os modelos *Donut*, *LayoutLM* e *Amazon Textract*, verificou-se que a solução da Amazon proporcionou maior precisão e eficiência na extração de informações, além de menor tempo de resposta. Embora o *Amazon Textract* envolva custos operacionais, sua viabilidade comercial e alto custo-benefício justificam sua escolha para o presente estudo.

4.1 Metodologia de Extração de Informações

O método proposto neste trabalho utilizará o serviço de consultas personalizadas do *Amazon Textract*, que fornece um modelo pré-treinado para a extração de dados de documentos através de perguntas e respostas (QA).

Com o objetivo de alcançar maior precisão e confiabilidade é realizado neste trabalho o treinamento deste modelo através de imagens de notas fiscais de energia elétrica, realizando o ajuste Fino, (*Fine-tuning*) do modelo nos dados dos conjuntos de treinamento propostos, adaptando-os para a tarefa específica de extração de informações das notas fiscais de energia elétrica para cada uma das distribuidoras de energia selecionadas para o presente trabalho. Desta forma são criados os adaptadores, modelos adaptados especificamente para cada uma das distribuidoras.

Dado a classificação da distribuidora de energia, processo citado em capítulo anterior e parte do fluxo de processamento geral proposto, é necessário selecionar as perguntas e o adaptador correto para a extração dos dados. No fluxograma descrito pela Figura 23, é possível observar este processo.

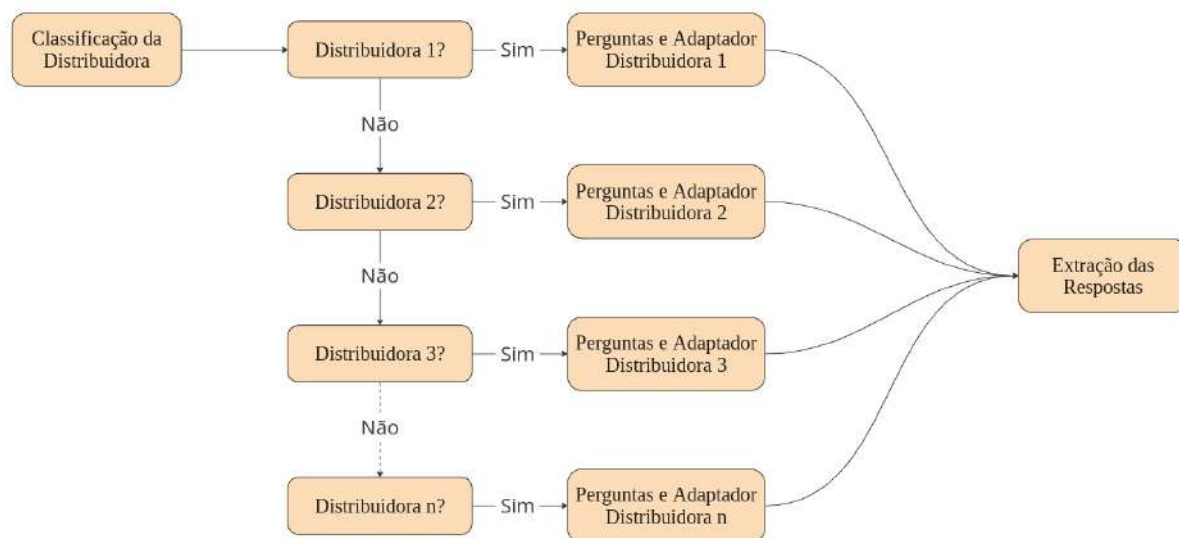


Figura 23 – Fluxograma de extração de informações.

Após o modelo retornar as respostas às perguntas enviadas ao adaptador, é necessário realizar um processamento adicional para garantir que as respostas estejam no formato exigido pelo software empresarial onde serão utilizadas. Na Figura 24, é apresentado o fluxograma do processo de pós-processamento, que termina na geração de um arquivo *JSON* contendo as informações organizadas no formato de chave-valor.



Figura 24 – Fluxograma de processamento das respostas.

4.1.1 Treinamento dos Modelos

O processo para treinar e otimizar um adaptador segue seis etapas principais, que podem ser observadas na Figura 25.



Figura 25 – Fluxograma de treinamento do adaptador (AMAZON WEB SERVICES, 2025).

- **Crie o Conjunto de Dados.**
O primeiro passo é construir um conjunto de dados, que é uma coleção de documentos e rótulos. Estes são utilizados para treinar e testar o adaptador. Uma vez criado o conjunto de dados e dividido entre conjuntos de treino e testes, ele estará pronto para ser usado.
- **Crie Consultas.**
Especificar as informações desejada. Para isso, devem ser utilizadas perguntas, que descrevem os dados de interesse de forma clara e direta. Logo, nesta etapa deve ser definida quais perguntas ou informações devem ser extraídas do conjunto de dados. As perguntas são utilizadas para realizar a extração das informações depois do treinamento do adaptador. Na Figura 27, é possível visualizar algumas das consultas criadas para o adaptador da concessionária *cemig*.
- **Verifique os Documentos.**
Nesta etapa, os documentos são analisados para garantir que os rótulos atendam aos requisitos de extração definidos. Nele os rótulos são revisados e ajustados para que correspondam exatamente às informações desejadas. Este processo pode ser visualizado na Figura 27.

- Treine o Adaptador.
Com os dados preparados, é treinado uma versão do adaptador, de forma a ensinar ao adaptador a extrair as informações com precisão.
 - Verifique as Métricas de Desempenho.
Após o treinamento, nesta etapa são analisadas as métricas de desempenho do adaptador, com base em métricas como precisão, recall e F1-score.
 - Precisão - ‘A precisão mede a porcentagem de informações extraídas (previsões) que estão corretas. Quanto maior a classificação de precisão, menos falsos positivos há’(AMAZON WEB SERVICES, 2025).
 - Recall - ‘O recall mede a porcentagem do total de itens relevantes que são identificados e extraídos com sucesso pelo modelo. Quanto maior o valor do recall, menos falsos negativos há’(AMAZON WEB SERVICES, 2025).
 - Pontuação F1 - ‘A pontuação F1 combina precisão e recall em uma única métrica, fornecendo uma medição equilibrada para precisão geral de extração’(AMAZON WEB SERVICES, 2025).
- Isso ajuda a determinar se ele precisa de treinamento adicional antes de ser utilizado. Uma ilustração do processo pode ser observado na Figura 26.
- Melhore o Adaptador.
Por fim, se necessário, a última etapa permite realizar modificações no conjunto de dados e realizar novos ciclos de treinamento para aprimorar a precisão do adaptador.

4.1.2 Validação e Pós-Processamento das Respostas

Embora os adaptadores apresentem alta precisão, erros ocasionais podem ocorrer devido a interpretações equivocadas ou "alucinações" dos modelos. Para mitigar esse



Figura 26 – Exemplo de métricas do modelo treinado fornecido pela interface (AMAZON WEB SERVICES, 2025).

1 páginas

Itens da Fatura	Unid.	Quant.	Preço Unit	Valor (R\$)	PIS/COF
Energia Elétrica	kWh	100	0,95254203	95,24	
Energia SCEE s/ ICMS	kWh	0,50816475	83,83		
Energia compensada GD I	kWh	0,48733000	-80,40		
Contrib Ilum Publica Municipal			84,29		
TOTAL				122,93	

Analisar as respostas [Práticas recomendadas para anotações](#)

O Textract rotulou os documentos para você como ponto de partida. Analise as respostas abaixo e edite-as de acordo com seus requisitos

Preço unit e valor da energia elétrica

+ Adicionar uma resposta

Quantidade preço unit e valor da energia s/ ICMS

+ Adicionar uma resposta

Quantidade preço unit e valor da energia compensada GD I

+ Adicionar uma resposta

Quer corrigir todos os resultados desta página? Limpe os resultados para remover todas as anotações prévias desta página e digite as respostas corretas.

[Limpar resultados](#)

[Cancelar](#) [Anterior](#) [Enviar e fechar](#)

Figura 27 – Processo de rotulação dos dados (AMAZON WEB SERVICES, 2025).

problema, propõe-se um mecanismo de validação que assegura que apenas respostas corroboradas pelo texto extraído sejam consideradas válidas.

O *Amazon Textract* retorna não apenas as respostas às consultas, mas também o texto integral da imagem. Assim, a validação é realizada por meio da verificação de expressões, garantindo que todas as palavras de uma resposta estejam presentes no texto extraído, como descrito no Algoritmo 3.

O pós-processamento complementa essa validação ao converter formatos e remover

Algoritmo 3 Validação de Expressões em Texto

Entrada: Expressão a ser validada, texto de referência

Saída: Resultado da validação da expressão

- 1: Dividir a expressão em palavras.
 - 2: **for** cada palavra na expressão **do**
 - 3: **if** a palavra não for encontrada no texto **then**
 - 4: **Retorne** "Expressão inválida".
 - 5: **end if**
 - 6: **end for**
 - 7: **Retorne** "Expressão válida".
-

inconsistências, garantindo que valores monetários e datas sejam extraídos corretamente. Por exemplo, expressões como "R\$ 286,17" são transformadas em "286.17" para integração com sistemas automatizados.

4.2 Resultados e Discussão

Os modelos foram treinados com imagens de faturas de energia elétrica, fornecida por *Digital Grid*, empresa parceira no desenvolvimento do projeto. As imagens utilizadas no treinamento passaram pela etapa de processamento de imagens propostas neste trabalho. As consultas criadas atendem a necessidade de extração de informações da empresa parceira, de forma que utilize os dados que os modelos estão capacitados a extrair em seu sistema.

Neste trabalho, foram treinados 9 adaptadores de 8 diferentes concessionárias de energia elétrica e com variadas consultas que extraem informações de consumo, energia compensada, taxa de iluminação, alíquotas, tarifas, datas de leitura, identificação da unidade consumidora, entre outras.

Na Tabela 9, é possível verificar a quantidade de faturas utilizadas no treinamento de cada um dos adaptadores. Os dados foram separados de forma a manter no conjunto de testes aproximadamente 30% do valor total de amostras obtidas e 70% restante no conjunto de treinamento.

Na Tabela 10, é possível as observar as métricas obtidas para cada um dos adaptadores treinados. É importante destacar que foram utilizados números de faturas diferentes para o treinamento de cada um dos adaptadores, o que implica em um número de faturas diferentes também no conjunto de testes para cada um dos adaptadores avaliados.

Tabela 9 – Número de faturas utilizados no treinamento em cada modelo.

Adaptador	Treinamento	Testes	Total
cemig 1	323	123	446
cemig 2	25	10	35
energisa	171	71	242
cpfl	84	37	121
Elektro	28	12	40
dme	18	8	26
enel ce	51	22	73
equatorial go 8	35	15	50
equatorial ma	38	16	54

O processo de treinamento se destaca pela praticidade em criar consultas e rotular imagens diretamente pela interface, tornando a tarefa intuitiva e eficiente. Além disso, o treinamento não exige que os desenvolvedores disponham de máquinas com alta capaci-

Tabela 10 – Resultados de Precisão, Recall e F1-score para diferentes distribuidoras em cada respectivo conjunto de teste.

Adaptador	Pontuação na F1	Precisão	Recall
cemig 1	0.973	0.973	0.973
cemig 2	0.964	0.964	0.964
energisa	0.988	0.988	0.987
cpfl	0.966	0.968	0.964
Elektro	0.978	0.978	0.978
dme	0.938	0.936	0.941
enel ce	0.849	0.849	0.849
equatorial go	0.848	0.856	0.839
equatorial ma	0.86	0.88	0.84

dade computacional, uma vez que a adaptação dos modelos é realizada remotamente nos servidores da fornecedora do sistema. A mesma lógica se aplica ao processo de inferência, ou seja, o modelo não é executado localmente, sendo necessário apenas uma conexão de rede estável para seu uso.

Entretanto, uma limitação relevante é o máximo de 30 consultas por adaptador no processo de treinamento, o que inviabiliza a criação de um adaptador único para atender todas as distribuidoras, ou de um único adaptador para uma única distribuidora, quando uma mesma distribuidora tem muitos casos diferentes e muitas informações a se extrair. Para a concessionária *cemig*, por exemplo, foi necessário desenvolver um segundo adaptador para abranger todos os casos existentes.

Apesar disso, os modelos treinados demonstram capacidade de extrair informações além das consultas originalmente configuradas. No entanto, para alcançar níveis mais altos de precisão e confiabilidade, pode ser necessário criar novas consultas sempre que surgem necessidades diferentes. Esse processo pode se tornar custoso em cenários dinâmicos, onde a variedade e a frequência de novas demandas de extração desafiam a escalabilidade do sistema.

Além disso, conforme descrito na documentação da *Amazon Textract*, a extração de pares de chave-valor em formulários funciona melhor quando a chave é utilizada diretamente como pergunta. Para outros tipos de extração, a recomendação é formular perguntas mais completas ([AMAZON WEB SERVICES, 2025](#)).

No entanto, em casos como a obtenção da linha do código de pagamento de uma fatura, não há uma chave ou termo explícito que identifique esse campo, o que compromete a precisão da extração. Esse padrão se repete em todos os adaptadores, tornando essa informação uma das que apresentam os piores resultados entre as consultas realizadas para cada um deles.

Para o adaptador *enel ce*, foram utilizadas imagens com qualidade inferior devido

a limitação na disponibilidade de imagens e observa-se a diminuição das métricas do adaptador quando comparado aos outros modelos com número próximo de faturas utilizadas no treinamento. Este resultado pode estar relacionada tanto à qualidade das imagens quanto às variações no layout do documento.

Já quando comparado a extração de informações dos adaptadores treinados com o modelo original sem o treinamento é notável a diferença na precisão da extração das informações. Para a fatura da Figura 22, com o adaptador *cemig 1*, o sistema é capaz de extrair todas as informações corretamente, enquanto o modelo original sem o ajuste fino, apresentou dificuldades na extração das informações utilizando as mesmas perguntas.

Na Tabela 11, é possível observar os dados extraídos, em verde estão os dados extraídos corretamente e em vermelho os dados extraídos incorretamente. Nota-se na tabela que mesmo sem o treinamento do adaptador o modelo pré-treinado é capaz de extrair informações corretamente, porém em situações que a informação está disposta no formato de tabela ou em que não há uma palavra chave claramente associada o modelo apresentou dificuldade, diferente do adaptador que consegue identificar as informações.

Vale destacar que parte dos dados sensíveis presentes na imagens também estão sensurados na Tabela 11.

Outro ponto importante é o tempo necessário para extração das consultas. Do *dataset*, proposto no Capítulo 3, foram escolhidos aleatoriamente 30 faturas referentes a distribuidora *cemig*. Além de verificar o tempo necessário para a extração, foi verificado no experimento o tempo médio para todas as etapas de processamento propostos. Na Tabela 12, é possível verificar o tempo médio de execução das etapas de processamento das faturas.

O experimento registra um tempo médio de execução do algoritmo de 35,273 segundos. Em comparação, a extração manual exige esse mesmo tempo apenas para obter o código de pagamento. Isso evidencia a capacidade do sistema de automatizar uma tarefa complexa e exaustiva quando realizada manualmente.

Vale destacar que, atualmente, a classificação do documento por texto é realizada apenas para a distribuidora *cemig*, pois é a única que possui mais de um adaptador. Isso significa que, para as demais distribuidoras, esse tempo não é contabilizado no processo de extração, a menos que haja a necessidade de múltiplos adaptadores no futuro.

Observa-se que o tempo necessário para a extração das consultas supera a soma de todas as outras etapas, tornando essa fase o principal gargalo de processamento. Ademais, como a execução em tempo real não é um requisito, o tempo dedicado à melhoria da qualidade da imagem, à análise de qualidade e à classificação do documento é considerado adequado para a aplicação.

Uma vez que o processamento ocorre de forma remota, o tempo necessário para

Tabela 11 – Comparação de dados de energia com adaptador treinado e sem treinamento.

Campo	Adaptador cemig 1	Sem Treinamento
Classe	Comercial Bifásico	Comercial Bifásico
Subclasse	Comercial	Comercial
Modalidade	Convencional B3	Convencional B3
Atual	06/12	-
Próxima	07/01	-
Nº do Cliente	72*****48	72*****48
Nº da Instalação	30*****01	30*****01
Total a pagar	92.12	92.12
Referente a	DEZ/2024	DEZ/2024
Vencimento	22/12/2024	22/12/2024
Emissão	06/12/2024	06/12/2024
Iluminação	34.52	34.52
Saldo de Crédito	446.27	446.27
Crédito Horário de Pico	-	3.328
Crédito Noturno	-	21
Consumo (kWh)	401	401
kWh COM IMPOSTOS	1.02651612	-
Custo de Energia	51.30	-
Energia Injetada (kWh)	351	-
kWh SEM ICMS	0.56390002	-
Valor Energia Compensada	197.91	-
Energia GDI	351	-
kWh SEM IMPOSTOS	0.54591161	-
Valor GD I	-191.61	-
Tarifa Unitária	0.54591161	0.54591161
Código da Distribuidora	8***00-6 9***01-8 6***11-8 08**38-1	0***38

a extração das consultas, não está sob o controle direto do desenvolvedor. Além disso, o número de consultas realizadas impacta diretamente o tempo total de resposta. Assim, a otimização desse processo passa, sobretudo, por uma adequada conexão com a rede e pela definição eficiente das consultas enviadas ao sistema.

As demais etapas do processamento, envolvem o uso de técnicas de visão computacional, redes neurais, aprendizado profundo, OCR e outras operações sobre imagens, e podem ter seu tempo reduzido com o uso de GPUs. No entanto, como o computador utilizado no experimento conta apenas com CPU, o desempenho dessas operações fica limitado à sua capacidade. Assim, uso de GPUs possibilitaria uma aceleração significativa

Tabela 12 – Tempo médio de execução do sistema de leitura.

Etapa	Tempo médio(s)
Detecção e Correção de Perspectiva do Documento	1,458
Remoção de Ruídos	0,713
Detecção e Correção de Orientação do Documento	0,180
Análise de Qualidade da Imagem	4,736
Classificação dos documentos por imagem	1,632
Classificação dos documentos por texto	1,589
Extração das Informações	24,645
Pós-processamento	0,003
Total	35,273

dessas tarefas, especialmente na inferência dos modelos utilizados.

Durante a execução do experimento, foi analisada a confiança média da melhor fatura retornada pelo processo de análise de qualidade. Observou-se um percentual médio de confiança de 66,84%, indicando que a maioria das faturas apresentava uma qualidade considerada ideal para a extração das informações.

Outra métrica analisada no experimento foi a acurácia. No caso da distribuidora *cemig*, foram extraídas 62 informações de cada fatura. Além disso, o experimento utilizou apenas o adaptador *cemig 1*, uma vez que nenhuma das faturas exigiu o uso do adaptador *cemig 2* ou das informações por ele extraídas.

Vale destacar que o método de classificação por texto é utilizado no sistema apenas quando o classificador por imagem detecta uma fatura *cemig*. A classificação por texto é utilizada para classificar faturas de subclasse *Irrigação Noturna*, com energia no horário reservado (HR). Caso a fatura pertença a essa subclasse o adaptador *cemig 2* é utilizado.

Com isso, o experimento alcançou uma acurácia de 99,67%, ou seja, em 99,67% dos dados extraídos, o sistema retornou a resposta correta, já também no formato correto, demonstrando sua viabilidade para uso em uma aplicação comercial de forma confiável e eficiente.

A maior parte das respostas incorretas ocorreram em consultas sobre informações de faturas do grupo tarifário A, que foram erroneamente detectadas em faturas do grupo B de subclasse baixa renda. Embora esse erro não comprometa a utilização das informações, ele ressalta a necessidade de ampliar o conjunto de treinamento do adaptador, incluindo um maior número de faturas do grupo A e da subclasse baixa renda, a fim de aprimorar a generalização do modelo.

5 Considerações Finais

Este trabalho investigou a viabilidade da automação do processamento de imagens e extração de dados de notas fiscais de energia elétrica, visando aprimorar a eficiência e a confiabilidade na gestão de créditos de energia. Para isso, foram desenvolvidas e avaliadas diversas técnicas de Visão Computacional e Inteligência Artificial, aplicadas desde o pré-processamento das imagens até a extração estruturada das informações contidas nos documentos.

Em outras palavras, este trabalho abordou a problemática do processamento automatizado de imagens de documentos. Esse desafio envolveu a correção da perspectiva e orientação de imagens de documentos, redução do ruído presente, classificação dos documentos, análise da qualidade das imagens e a extração automática de informações.

Dessa forma, com o objetivo de solucionar o corrigir a perspectiva e realizar a detecção do documento, o modelo de segmentação de instâncias treinado foi capaz de segmentar o documento presente na imagem, com precisão de 0,997. Além disso, com o algoritmo para detecção de vértices e correção de perspectiva, a técnica proposta foi capaz de remover completamente o fundo da imagem e grande parte da distorção de perspectiva.

Ademais, os filtros utilizados foram capazes de reduzir parcialmente o ruído presente na imagem, e o algoritmo de detecção e correção de orientação, que utiliza o método estatístico *Z-Score*, é capaz de determinar a orientação do documento com um erro médio de 0,10795 graus no *dataset* proposto para avaliação, eliminando assim distorções de orientação do documento com precisão.

Os resultados sugerem que os algoritmos empregados contribuíram para uma melhora na qualidade das imagens. Dessa forma, os algoritmos propostos podem ser aplicados às etapas de processamento de imagens de documentos em projetos de análise e extração de dados automatizada.

O desenvolvimento desta ferramenta resultou na detecção automática de notas fiscais de energia elétrica de forma eficiente, com a qualidade das imagens das notas fiscais aprimorada, com a diminuição de ruídos e distorções geométricas, tornando as informações mais nítidas e legíveis, garantindo que os dados estejam sempre dispostos de maneira adequada para facilitar a extração.

O método de análise da qualidade da imagem por meio da análise de confiança das caixas de texto permitiu selecionar a melhor imagem para subsequente análise, corrigindo possíveis falhas da etapa de processamento do documento e impedindo a análise de documentos que não possuem qualidade suficiente e, conseqüentemente, dados imprecisos.

Já o classificador de *layouts* de faturas de energia elétrica tornou o projeto mais flexível e adaptado para cada uma das distribuidoras de energia, selecionando corretamente o modelo e as informações a serem extraídas.

Além disso, os modelos treinados com o *Amazon Textract* se mostraram eficientes na extração de informações específicas das imagens, capturando informações com alta precisão. Os métodos de validação propostos aumentam sua confiabilidade e precisão, minimizando erros e tornando a extração de dados mais confiável, e os métodos de pós-processamento facilitam a integração da ferramenta a um software empresarial.

Apesar dos avanços obtidos neste trabalho, são observados lacunas a serem exploradas e pontos a serem melhorados. O principal ponto a ser destacado é a necessidade de maior dinamismo e generalização dos modelos treinados. Mesmo que a interface seja intuitiva e facilite o processo de rotulação dos dados, a necessidade contínua de extração de novas informações das faturas de energia elétrica exige modelos e métodos com maior capacidade de generalização.

Logo, em relação a pergunta de pesquisa e hipóteses propostas. A hipótese **H1**, que questionava se a extração de dados poderia ser automatizada com alta precisão utilizando técnicas de Inteligência Artificial e Visão Computacional, foi confirmada. O sistema desenvolvido foi capaz de melhorar a qualidade dos documentos com alta precisão através de diferentes técnicas de Visão Computacional e extrair informações relevantes utilizando o *Amazon Textract*, demonstrando eficiência na captura dos dados. Além disso, os métodos de validação e pós-processamento propostos aumentaram a confiabilidade da extração, minimizando erros.

A segunda hipótese **H2**, que investigava se a melhoria na qualidade das imagens influenciaria na acurácia da extração de dados, também foi validada. A remoção de ruídos e distorções geométricas aumentou significativamente a média de confiança dos textos extraídos dos documentos através do *Tesseract OCR*, permitindo uma extração mais precisa.

Quanto à terceira hipótese **H3**, que avaliava se a ferramenta automatizada poderia substituir o processamento manual, os resultados indicaram que a solução proposta é mais eficiente. A automação da extração reduziu a necessidade de intervenção humana e mitigou erros associados ao processamento manual, tornando o sistema mais confiável e ágil. Entretanto, observou-se que o tempo de processamento ainda pode ser otimizado, sendo esse um ponto a ser aprimorado em trabalhos futuros.

Por fim, a quarta hipótese **H4**, que questionava a generalização dos métodos para diferentes formatos e padrões de notas fiscais, não pôde ser plenamente validada. Embora o classificador de *layouts* tenha permitido a adaptação do sistema a diversos *layouts* conhecidos de distribuidoras de energia, essa adaptação depende do conhecimento prévio desses *layouts*. O sistema demonstrou flexibilidade para lidar com formatos previamente

treinados, mas o sistema não generaliza de forma ampla, atuando como um especialista que requer conhecimento prévio para cada novo padrão.

Dessa forma, este estudo demonstrou a viabilidade e a relevância do uso da IA e Visão Computacional para a extração automatizada de informações de notas fiscais de energia elétrica. Ademais, trabalhos futuros devem explorar novas abordagens e métodos de extração de informações que reduzam o tempo de extração dos dados, uma vez que essa etapa representou a maior demanda de tempo no sistema. Dessa forma, otimizar o desempenho geral do sistema depende diretamente da melhoria desse aspecto.

Logo, a integração bem-sucedida da ferramenta em um software empresarial irá permitir sua utilização direta nos fluxos de trabalho. A ferramenta contribuirá significativamente para a gestão eficiente do compartilhamento de crédito de energia entre proprietários de sistemas de geração de energia renovável e consumidores, otimizando o processo e reduzindo erros manuais.

Além disso, o impacto do projeto é evidente não apenas na gestão eficiente do compartilhamento de energia, mas também em sua aplicabilidade a outros contextos que envolvem o processamento de grandes volumes de documentos, como faturas, contratos e recibos. Com isso, a solução proposta promove a economia de recursos humanos, aumenta a produtividade e reduz a ocorrência de erros associados ao processamento manual.

Os resultados e as inovações obtidas no projeto foram sistematicamente registrado neste documento e organizados para a elaboração e publicação de artigos científicos. Esses artigos foram submetidos a conferências e revistas, contribuindo para o avanço nas áreas de visão computacional e automação de processos baseados em documentos, incentivando a troca de conhecimento e novas pesquisas na área.

Ademais, este trabalho se insere em um campo de pesquisa inovador ao abordar desafios complexos, como a extração de informações estruturadas de documentos despadronizados e a integração de métodos avançados de Inteligência Artificial e Visão Computacional. O projeto contribui diretamente para a modernização da gestão de energia e reforça a importância de soluções tecnológicas no suporte a iniciativas sustentáveis e eficientes.

Referências

- ABBAS, S. et al. Convolutional neural network based intelligent handwritten document recognition. **Computers, Materials & Continua**, v. 70, n. 3, 2022.
- AHMAD, R.; NAZ, S.; RAZZAK, I. Efficient skew detection and correction in scanned document images through clustering of probabilistic hough transforms. **Pattern recognition letters**, Elsevier, v. 152, p. 93–99, 2021.
- AKHTER, S. S. M. N.; REGE, P. P. Improving skew detection and correction in different document images using a deep learning approach. In: IEEE. **2020 11th International Conference on Computing, Communication and Networking Technologies (ICCCNT)**. Kharagpur, India, 2020. p. 1–6.
- AL-SHATNAWI, A. M.; OMAR, K. Skew detection and correction technique for arabic document images based on centre of gravity. **Journal of Computer Science**, v. 5, n. 5, p. 363, 2009.
- AMAZON WEB SERVICES. **Evaluating and Improving Amazon Textract Adapter Models**. 2025. Disponível em: <https://docs.aws.amazon.com/en_us/textract/latest/dg/textract-evaluating-improving-adapters.html>. Acesso em: 26 jan. 2025.
- ANEEL. **Geração Distribuída**. 2025. Disponível em: <<https://www.gov.br/aneel/p-t-br/assuntos/geracao-distribuida>>. Acesso em: 6 fev. 2025.
- ATTIVISSIMO, F. et al. An automatic reader of identity documents. In: IEEE. **2019 IEEE International Conference on Systems, Man and Cybernetics (SMC)**. Bari, Italy, 2019. p. 3525–3530.
- BAJRAMI, M. et al. A comprehensive analysis of layoutlm and donut for document classification. Ss Cyril and Methodius University in Skopje, Faculty of Computer Science and . . . , 2023.
- BAVISKAR, D. et al. Efficient automated processing of the unstructured documents using artificial intelligence: A systematic literature review and future directions. **IEEE Access**, v. 9, p. 72894–72936, 2021.
- BHATT, D. et al. Cnn variants for computer vision: History, architecture, application, challenges and future scope. **Electronics**, MDPI, v. 10, n. 20, p. 2470, 2021.
- BOIANGIU, C.-A. et al. Voting-based document image skew detection. **Applied Sciences**, MDPI, v. 10, n. 7, p. 2236, 2020.
- BRADSKI, G. The OpenCV Library. **Dr. Dobb's Journal of Software Tools**, 2000.
- CASTELBLANCO, A. et al. Machine learning techniques for identity document verification in uncontrolled environments: A case study. In: SPRINGER. **Pattern Recognition: 12th Mexican Conference, MCPR 2020, Morelia, Mexico, June 24–27, 2020, Proceedings 12**. Morelia, Mexico, 2020. p. 271–281.

DEEPAK, S.; AMEER, P. Automated categorization of brain tumor from mri using cnn features and svm. **Journal of Ambient Intelligence and Humanized Computing**, Springer, v. 12, n. 8, p. 8357–8369, 2021.

DIGITAL GRID. **Gestão de Créditos de Energia na Geração Distribuída**. 2025. Disponível em: <<https://dg.energy/>>. Acesso em: 6 fev. 2025.

DING, Y. et al. V-doc: Visual questions answers with documents. In: IEEE. **Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)**. New Orleans, USA, 2022. p. 21492–21498.

EBRAHIMPOUR, N. Iris recognition using mobilenet for biometric authentication. In: IKSAD. **9th INTERNATIONAL ZEUGMA CONFERENCE ON SCIENTIFIC RESEARCH, Gaziantep, Turkey**. Gaziantep, Turkey, 2023.

GUÉRIN, J. et al. Combining pretrained cnn feature extractors to enhance clustering of complex natural images. **Neurocomputing**, Elsevier, v. 423, p. 551–571, 2021.

HA, H. T.; HORÁK, A. Information extraction from scanned invoice images using text analysis and layout features. **Signal Processing: Image Communication**, Elsevier, v. 102, p. 116601, 2022.

HARLEY, A. W.; UFKES, A.; DERPANIS, K. G. Evaluation of deep convolutional nets for document image classification and retrieval. In: IEEE. **2015 13th International Conference on Document Analysis and Recognition (ICDAR)**. Tunis, Tunisia, 2015. p. 991–995.

JAFARI, Z.; KARAMI, E. Breast cancer detection in mammography images: a cnn-based approach with feature selection. **Information**, MDPI, v. 14, n. 7, p. 410, 2023.

JIANG, S. et al. Deep learning for technical document classification. **IEEE Transactions on Engineering Management**, IEEE, v. 71, p. 1163–1179, 2022.

JOCHER, G.; CHAURASIA, A.; QIU, J. **Ultralytics YOLOv8**. 2023. Disponível em: <<https://github.com/ultralytics/ultralytics>>.

KANG, L. et al. Convolutional neural networks for document image classification. In: IEEE. **2014 22nd international conference on pattern recognition**. Stockholm, Sweden, 2014. p. 3168–3172.

KANG, L. et al. Multi-page document visual question answering using self-attention scoring mechanism. In: SPRINGER. **International Conference on Document Analysis and Recognition**. Athens, Greece, 2024. p. 219–232.

KIM, G. et al. Donut: Document understanding transformer without ocr. **arXiv preprint arXiv:2111.15664**, v. 7, n. 15, p. 2, 2021.

LIU, Y.; PU, H.; SUN, D.-W. Efficient extraction of deep image features using convolutional neural network (cnn) for applications in detecting and analysing complex food matrices. **Trends in Food Science & Technology**, Elsevier, v. 113, p. 193–204, 2021.

- LUNG. Open Source Dataset, **document-segmentation-v2 Dataset**. Roboflow, 2024. <<https://universe.roboflow.com/lung-x8e11/document-segmentation-v2-gt86h>>. Disponível em: <<https://universe.roboflow.com/lung-x8e11/document-segmentation-v2-gt86h>>.
- MISHRA, A. et al. Ocr-vqa: Visual question answering by reading text in images. In: IEEE. **2019 International Conference on Document Analysis and Recognition (ICDAR)**. Sydney, Australia, 2019. p. 947–952.
- NASIR, I. M. et al. Pearson correlation-based feature selection for document classification using balanced training. **Sensors**, MDPI, v. 20, n. 23, p. 6793, 2020.
- NOOR, M. N. et al. Gastronet: A robust attention-based deep learning and cosine similarity feature selection framework for gastrointestinal disease classification from endoscopic images. **CAAI Transactions on Intelligence Technology**, Wiley Online Library, 2023.
- Projeto Solar. **Usina Geração Compartilhada**. 2025. Disponível em: <<https://www.projetosolar.eng.br/usina-geracao-compartilhada>>.
- RAMTEKE, R.; PATHAN, I. K.; MEHROTRA, S. Skew angle estimation of urdu document images: a moments based approach. **International Journal of Machine Learning and Computing**, IACSIT Press, v. 1, n. 1, p. 7, 2011.
- RIGO, P. D. et al. Competitive business model of photovoltaic solar energy installers in brazil. **Renewable Energy**, v. 181, p. 39–50, 2022. ISSN 0960-1481. Disponível em: <<https://www.sciencedirect.com/science/article/pii/S096014812101329X>>.
- SIDHWA, H. et al. Text extraction from bills and invoices. In: IEEE. **2018 international conference on advances in computing, communication control and networking (ICACCCN)**. Greater Noida, India, 2018. p. 564–568.
- SINGH, C.; BHATIA, N.; KAUR, A. Hough transform based fast skew detection and accurate skew correction methods. **Pattern Recognition**, Elsevier, v. 41, n. 12, p. 3528–3546, 2008.
- SRIVASTAVA, D.; HARIT, G. Simultaneous denoising and super resolution of document images. **Sādhanā**, Springer, v. 49, n. 1, p. 35, 2024.
- TAN, M.; LE, Q. Efficientnet: Rethinking model scaling for convolutional neural networks. In: PMLR. **International conference on machine learning**. Long Beach, USA, 2019. p. 6105–6114.
- TASCI, B. Automated ischemic acute infarction detection using pre-trained cnn models' deep features. **Biomedical Signal Processing and Control**, Elsevier, v. 82, p. 104603, 2023.
- TROPIN, D. V. et al. Approach for document detection by contours and contrasts. In: IEEE. **2020 25th International Conference on Pattern Recognition (ICPR)**. Milan, Italy, 2021. p. 9689–9695.
- XIA, P.; ZHANG, L.; LI, F. Learning similarity with cosine similarity ensemble. **Information sciences**, Elsevier, v. 307, p. 39–52, 2015.

XU, Y. et al. Layoutlm: Pre-training of text and layout for document image understanding. In: ACM. **Proceedings of the 26th ACM SIGKDD International Conference on Knowledge Discovery & Data Mining**. Virtual Event, USA, 2020. p. 1192–1200.

YE, J. Cosine similarity measures for intuitionistic fuzzy sets and their applications. **Mathematical and computer modelling**, Elsevier, v. 53, n. 1-2, p. 91–97, 2011.