Projeto Final

Esteganografia

Segurança da Informação

Gustavo Silveira Ribeiro - 156.560
Laura Mª Cunha Lisbôa - 163.882

São José dos Campos
2025

1 Resumo

Este trabalho consiste no estudo da técnica de esteganografia e a comparação entre diferentes métodos, como *Least Significant Bit*, *Discrete Cosine Transform* e *Discrete Wavelet Transform*. Assim, será demonstrado neste artigo como foi realizada a implementação de esteganografia ao esconder mensagens em imagens; treinar os métodos para identificar imagens originais e alteradas; e decifrar o objeto que foi escondido nas imagens trabalhadas.

2 Introdução

A palavra derivada do grego "steganos"e "graphia"combinam as definições de grafia escondida. Assim, a esteganografia pode ser definida como "a arte de esconder informações". Com isso, trata-se de uma técnica que é usada para ocultar informações dentro de outros arquivos, de maneira que não é possível perceber.

Essa prática tem ganhado certa popularidade nos dias atuais, mas é algo que acontece desde muitos séculos passados. Os romanos, durante guerras, ao enviar mensagens, tinham o costume de esconder detalhes nas correspondências, pois havia a possibilidade de inimigos terem acesso às informações. Era uma precaução para que não fossem prejudicados com o vazamento de suas estratégias.

Dentro dessa ótica, atualmente, a esteganografia avançou muito, pois é possível esconder muito mais do que simples mensagens, mas também áudios e vídeos, de maneira que não seja aparente por alguém que observe o arquivo que possui o conteúdo ocultado.

À medida que desempenha um papel importante dentro da área de segurança da informação, a esteganografia também oferece preocupação por estar sendo utilizada cada vez mais por *hackers*, principalmente em operações de roubo de informações financeiras e espionagem virtual. Assim, é necessário ter atenção aos riscos oferecidos dentro do ramo da tecnologia, pois eles existem de inúmeras formas.

Portanto, faz-se um tema extremamente importante de ser abordado e entendido dentro de ambientes acadêmicos e corporativos, a fim de que as pessoas tenham maior ciência e possam estar sempre mais atentas às técnicas que, por mais que sejam aplicadas por profissionais de forma ética, também podem ser executadas com más intenções.

Este artigo demonstra a implementação da esteganografia para esconder mensagens em imagens e a comparação entre alguns métodos que podem ser implementados para esconder mensagens, extraí-las, além de treinar os métodos para identificar imagens originais e alteradas.

3 Revisão de literatura

Atualmente, a esteganografia é amplamente utilizada tanto para fins legítimos, como proteção de direitos autorais e marca d' água digital, quanto para intenções maliciosas, incluindo a transmissão oculta de informações em ataques cibernéticos.

3.1 Métodos de esteganografia

Os principais métodos de esteganografia digital analisados neste estudo são:

Least Significant Bit (**LSB**): Considerado um dos métodos mais simples e comuns, consiste em modificar os bits menos significativos dos pixels de uma imagem para armazenar a informação oculta. Essa técnica apresenta alta capacidade de armazenamento e pouca alteração visual, mas é vulnerável a compressão e análises estatísticas.

Discrete Cosine Transform (**DCT**): Utilizado principalmente em imagens comprimidas, como JPEG, este método insere informação nos coeficientes da transformação discreta do cosseno, tornando-a menos perceptível e mais resistente a compressão.

Discrete Wavelet Transform (**DWT**): Similar ao DCT, mas baseado na decomposição da imagem em diferentes níveis de frequência, permitindo uma maior robustez contra ataques e compressão.

Cada método tem suas vantagens e desvantagens, quando observados em relação à capacidade de armazenamento, robustez e impacto visual.

3.2 Aplicabilidades e desafios

A esteganografia tem aplicações em diversas áreas, incluindo:

- Proteção de direitos autorais em multimídia;
- Transmissão segura de mensagens em ambientes de risco;
- Uso por hackers para esconder malware em imagens e arquivos.

Dessa maneira, os principais desafios envolvem a detecção de conteúdo oculto por análises forenses, a resistência a compressão e edições, além das questões éticas relacionadas ao seu uso.

3.3 Trabalhos relacionados

Estudos recentes exploram métodos avançados de esteganografia e sua detecção. Por exemplo, pesquisas aplicando redes neurais para identificar padrões em imagens adulteradas demonstram avanços na análise forense digital. Além disso, novos algoritmos esteganográficos têm sido desenvolvidos para aumentar a segurança e a capacidade de ocultamento.

4 Metodologia

A metodologia deste trabalho é composta pelos seguintes passos:

- Estudo dos três métodos (Least Significant Bit, Discrete Cosine Transform e Discrete Wavelet Transform);
- Seleção de imagens;
- Seleção de mensagens;
- Testes.

4.1 Estudo dos três métodos

Inicialmente, para que o desenvolvimento fosse iniciado, foi necessário entender como procede cada um dos métodos. Dessa forma, algumas fontes de informação foram essenciais para que o projeto ocorresse como o esperado.

Dessa forma, os conceitos do primeiro método, o **LSB** (Least Significant Bit) consistem na técnica de esteganografia baseada na substituição dos bits menos significativos de um arquivo digital, como uma imagem ou um áudio, para ocultar informações. Como a alteração ocorre nos bits de menor peso, as mudanças visuais ou sonoras são quase imperceptíveis ao usuário comum. No entanto, esse método é vulnerável à compressão, pois padrões podem ser identificados em arquivos manipulados.

A **DCT** (Discrete Cosine Transform) é amplamente utilizada em compressão de imagens, como no formato JPEG, e também em esteganografia. Esse método divide a imagem em blocos e converte os valores espaciais em coeficientes de frequência, o que permite que a inserção de dados ocorra nas frequências médias sem comprometer significativamente a qualidade visual. Por ser integrada a formatos comprimidos, a esteganografia baseada em DCT oferece maior robustez contra compressão e transformações simples, mas pode ser detectada por análise de padrões em coeficientes modificados.

Por fim, a **DWT** (Discrete Wavelet Transform) funciona de forma semelhante à DCT, mas ao invés de dividir a imagem em blocos fixos, ela decompõe os dados em diferentes níveis de frequência por meio de funções wavelet. Esse processo permite manipular componentes de baixa e alta frequência separadamente, tornando a ocultação de dados mais eficiente e resistente à compressão e filtros de ruído. Como a DWT preserva melhor as características estruturais da imagem, é uma técnica mais robusta para aplicações esteganográficas em comparação com LSB e DCT.

4.2 Seleção de imagens e mensagens

Após a tentativa de manipular imagens de um dataset que possuía mais de 20000 imagens, foi decidido construir uma base de dados própria, a fim de ter mais autonomia sobre estilo das imagens e melhor manipulação delas em domínios pessoais (*Google Drive*, por exemplo).

As fotos foram todas retiradas do site https://unsplash.com/pt-br. Selecionadas a dedo, 450 imagens compuseram a base de dados. Ademais, para cada método, foi decidido escolher um tema para que fossem mais justos os testes. Como exemplo, em LSB, os temas escolhidos foram florestas e cidades, pois fornecem maior variedade de corese texturas. As imagens trabalhadas em DCT eram de praia, mais voltadas para a areia e água do mar, pois, assim, haveria mais frequência nos detalhes. E, por sua vez, em DWT, as fotos eram voltadas para montanhas, neve e rochas, na intenção de obter detalhes em múltiplas escalas.

Em relação ao conteúdo para esconder nas imagens, três arquivos no formato *.txt* foram organizados com diversos textos. O primeiro, utilizado no método LSB, continha passagens da bíblia; o segundo, para DCT, trechos do Código Penal; e, em sequência para DWT, trechos do livro "O pequeno príncipe", de Antoine de Saint-Exupéry.

4.3 Testes

Para realizar os testes, os códigos foram todos implementados no *Google Colab*, a fim de facilitar o processo de *coding* e análise das saídas.

Neste estudo, foram inseridas mensagens que tinham em torno de 70 bytes nas imagens utilizando três diferentes métodos de esteganografia: **LSB**, **DCT** e **DWT**. Dessa forma, cada técnica possui características distintas na forma como os dados são ocultados, o que influencia diretamente na qualidade visual da imagem resultante e sua resistência à detecção. Então, o objetivo foi avaliar a eficácia de cada método na ocultação de informações sem comprometer significativamente a integridade perceptiva das imagens.

Após a inserção, as mensagens foram extraídas das imagens para verificar a precisão e a fidelidade da recuperação das informações. Cada método apresentou diferentes níveis de sucesso na extração, dependendo da robustez da técnica utilizada e das possíveis alterações sofridas pelas imagens durante o processo. Esse teste foi essencial para validar a eficiência de cada abordagem, identificando quais técnicas são mais adequadas para aplicações práticas e quais são mais suscetíveis a perdas de dados.

Além da inserção e extração das mensagens, foram utilizados algoritmos de *machine learning*: **SVM** (Support Vector Machine) e **Random Forest**, para diferenciar imagens originais das modificadas. O SVM é um classificador supervisionado que busca encontrar um hiperplano ótimo para separar diferentes classes, sendo eficiente em problemas de alta dimensionalidade. Já o Random Forest é um conjunto de árvores de decisão que melhora a precisão da classificação ao combinar os resultados de múltiplos modelos. A aplicação dessas técnicas permitiu avaliar o nível de detecção das imagens modificadas, fornecendo insights sobre a discrição e a eficácia de cada abordagem esteganográfica diante de análise forense automatizada.

5 Análises

Nesta seção, discutem-se os resultados obtidos com os testes realizados, abordando tanto a eficácia dos métodos de inserção e extração de mensagens quanto a capacidade dos algoritmos de machine learning (SVM e Random Forest) em diferenciar imagens originais das alteradas.

5.1 Análise dos Resultados de Inserção e Extração

Primeiramente, foram inseridos aproximadamente 70 bytes de informação em cada imagem e, em seguida, a fim de realizar uma comparação, em torno de 300 bytes em mensagens, utilizando os métodos LSB, DCT e DWT, sendo os seguintes pontos observados:

LSB:

Qualidade Visual: As alterações nos bits menos significativos resultaram em modificações praticamente imperceptíveis à análise visual.

Robustez: Apesar da boa preservação estética, o método mostrou-se vulnerável a compressões e manipulações, o que pode ocasionar perda ou corrupção dos dados ocultos.

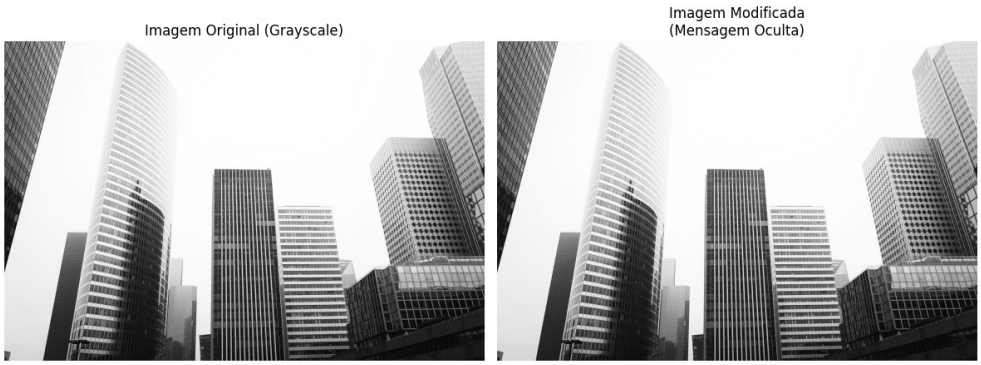

Figura 1: Imagem do dataset de LSB. Original e modificada lado a lado.

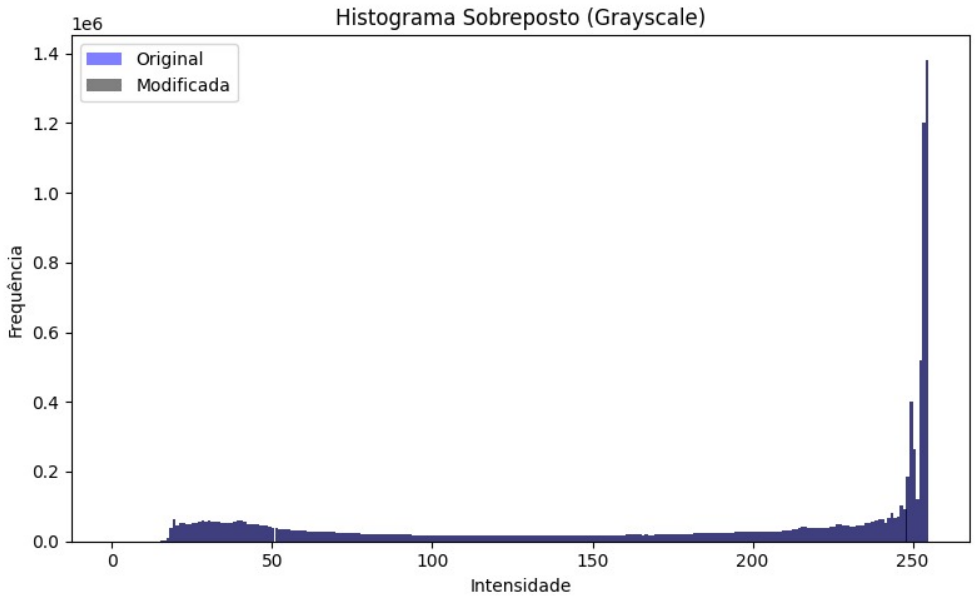

Figura 2: Histograma de LSB.

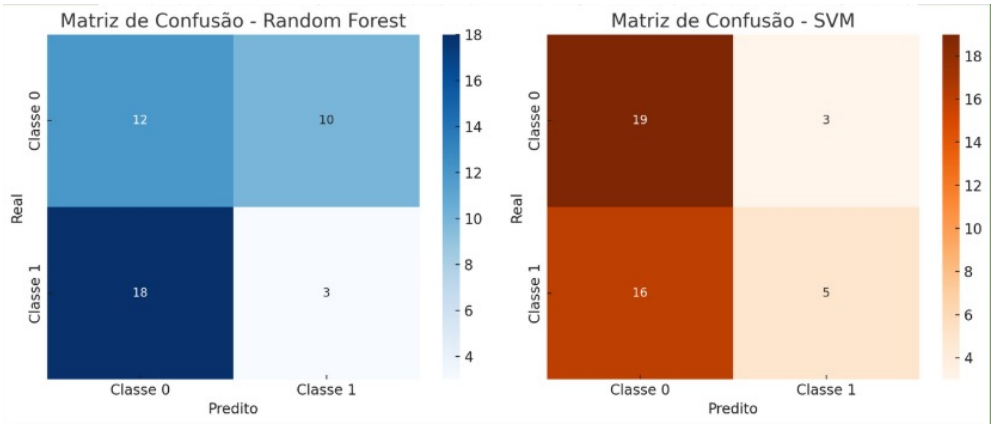

Figura 3: Matriz de confusão SVM e Random Forest para LSB. Mensagens com 70 bytes

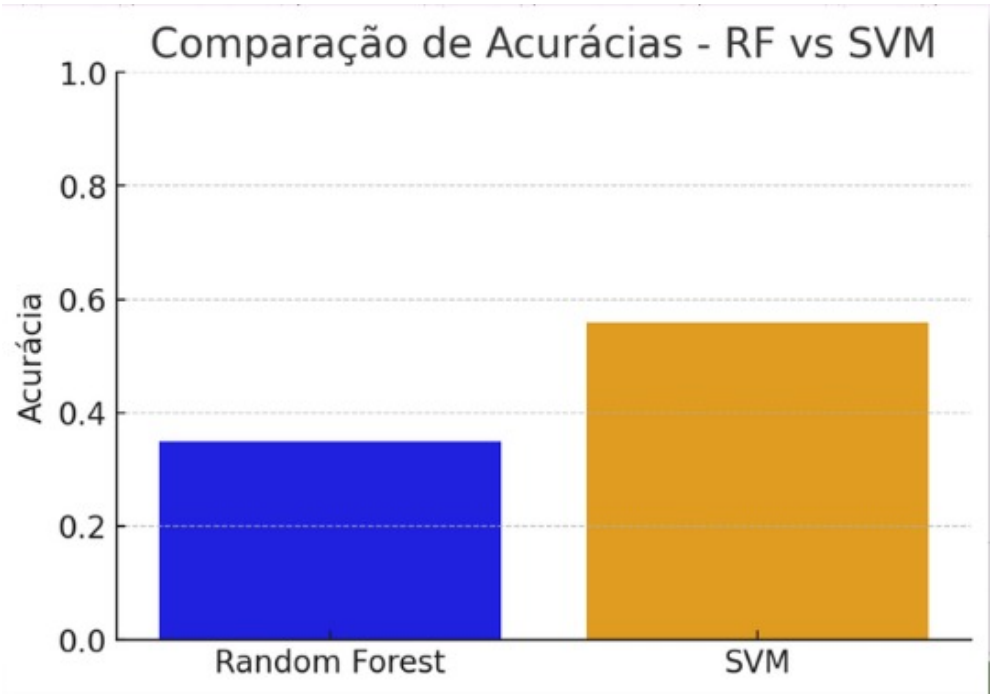

Figura 4: Acurácia SVM e Random Forest para LSB. Mensagens com 70 bytes.

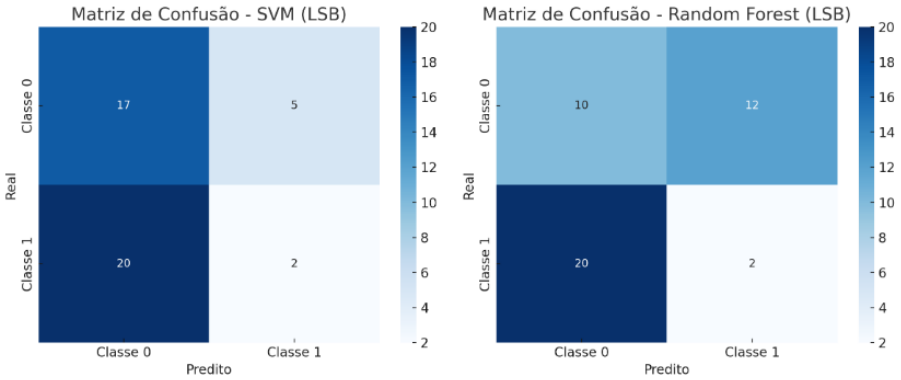

Figura 5: Matriz de confusão SVM e Random Forest para LSB. Mensagens com 300 bytes.

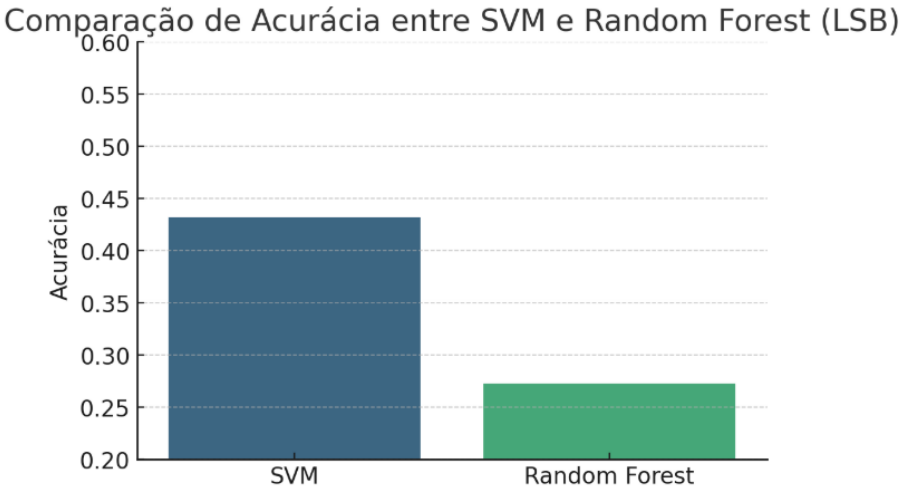

Figura 6: Acurácia SVM e Random Forest para LSB. Mensagens com 300 bytes.

Figura 7: Imagem do dataset de DCT. Original e modificada lado a lado.

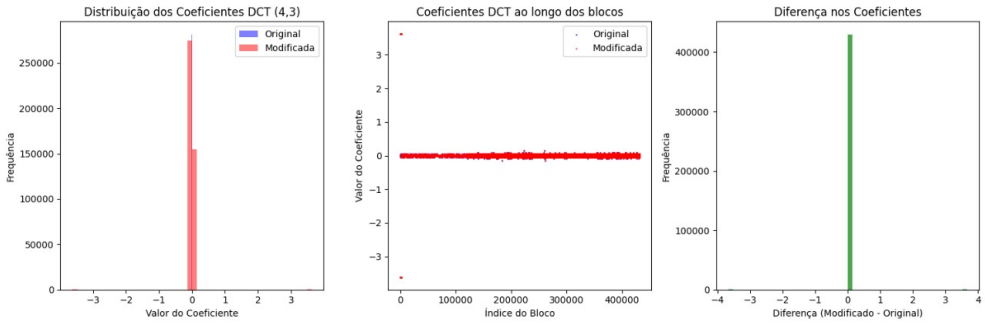

Figura 8: Visão gráfica dos coeficientes DCT.

Extração: A recuperação dos dados foi precisa quando as imagens não passaram por transformações, mas apresentou dificuldades quando submetidas a compressões ou edições.

DCT:

Qualidade Visual: A transformação dos blocos de imagem e a modificação dos coeficientes de frequência permitiram ocultar os dados sem comprometer significativamente a aparência da imagem, sobretudo em áreas de média frequência.

Robustez: Demonstrou maior resistência à compressão (ex.: JPEG), mas a alteração dos coeficientes pode ser detectada por análises estatísticas detalhadas.

Extração: Os testes indicaram uma taxa de extração satisfatória em condições controladas, embora a precisão dependa da integridade dos coeficientes após compressões ou filtros.

DWT:

Qualidade Visual: A decomposição em múltiplos níveis de frequência possibilitou uma inserção dos dados de forma mais "discreta", preservando as carac-

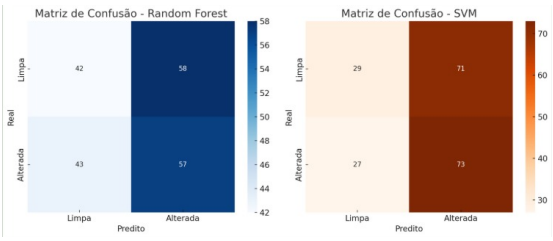

Figura 9: Matriz de confusão SVM e Random Forest para DCT. Mensagens com 70 bytes.

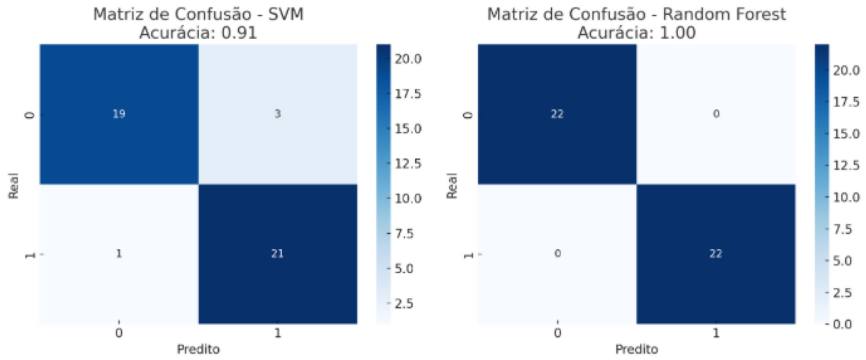

Figura 10: Matriz de confusão SVM e Random Forest para DCT. Mensagens com 300 bytes.

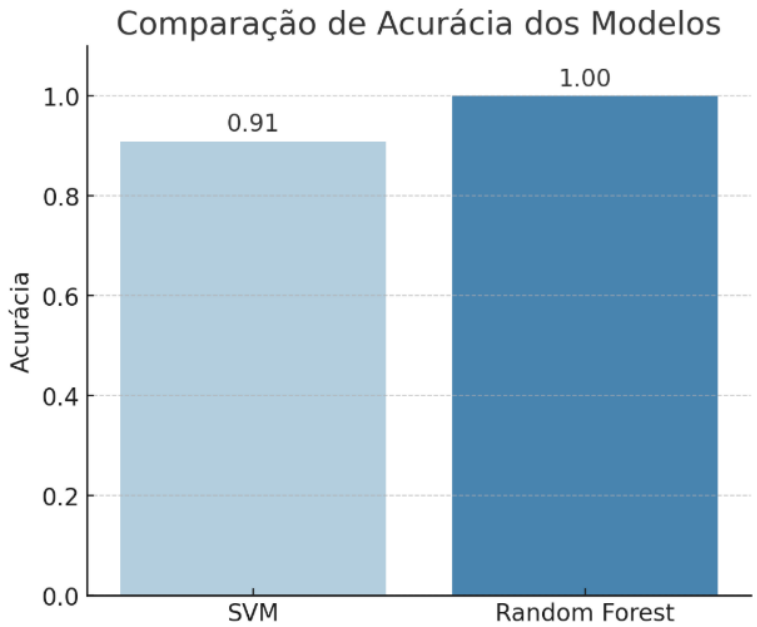

Figura 11: Acurácia SVM e Random Forest para DCT. Mensagens com 300 bytes.

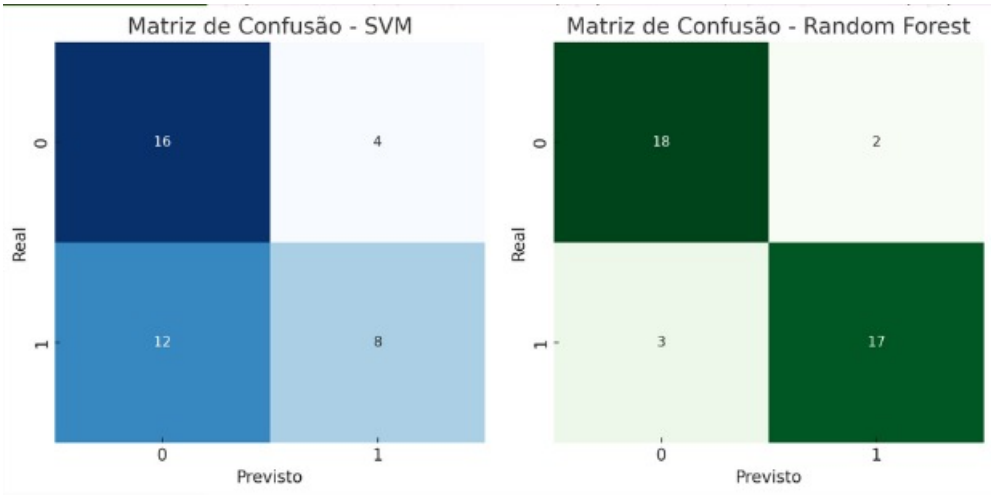

Figura 12: Matriz de confusão SVM e Random Forest para DWT. Mensagens com 70 bytes.

terísticas estruturais da imagem.

Robustez: Foi o método que apresentou maior resiliência frente a compressões e ataques por ruído, mantendo a integridade dos dados ocultos mesmo após modificações na imagem.

Extração: A extração mostrou-se consistente e com alta fidelidade, corroborando a eficiência da abordagem DWT em cenários com edições e compressões moderadas.

Além das análises qualitativas, foram calculadas métricas de qualidade (como PSNR – Peak Signal-to-Noise Ratio) para quantificar a degradação visual. Os valores de PSNR indicaram que, embora o método LSB apresente uma leve redução na qualidade comparado aos métodos DCT e DWT, as diferenças permanecem dentro de níveis aceitáveis para aplicações em que a alteração imperceptível é essencial.

5.2 Análise dos Resultados de Classificação

Utilizando os algoritmos de machine learning SVM e Random Forest, procedeu-se à identificação de imagens modificadas em relação às originais. As análises de classificação apontaram:

SVM:

Demonstrou bom desempenho na separação entre as classes (imagens originais e modificadas), com uma taxa de acurácia elevada em condições ideais. Constatou-se que alterações promovidas pelo método LSB eram mais facilmente detectáveis, dada a alteração direta dos bits, enquanto as alterações dos métodos DCT e DWT apresentaram padrões mais sutis.

Random Forest:

A combinação de múltiplas árvores de decisão possibilitou uma robusta classificação, atingindo resultados comparáveis aos do SVM. Observou-se uma maior resiliência do método frente a variações nos dados, o que sugere que técnicas ensemble podem ser mais adequadas em cenários onde as alterações esteganográficas são mínimas. Em ambos os casos, os testes evidenciaram que, embora seja possível detectar alterações introduzidas pelos métodos de esteganografia, a eficácia da detecção varia conforme a técnica utilizada para ocultamento. Enquanto o LSB

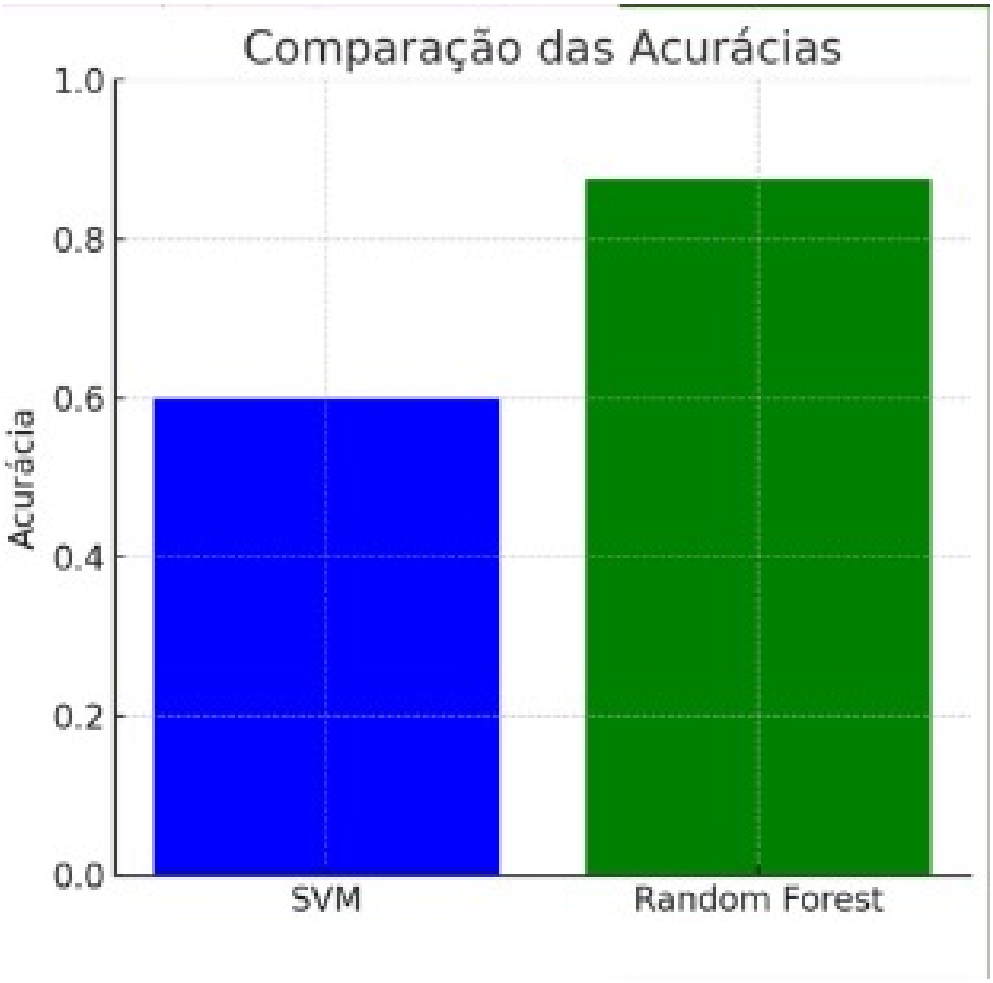

Figura 13: Acurácia SVM e Random Forest para DWT. Mensagens com 70 bytes.

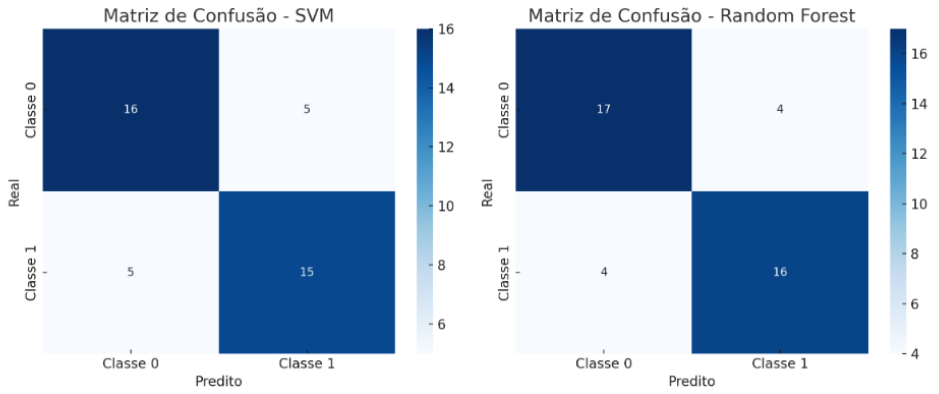

Figura 14: Matriz de confusão SVM e Random Forest para DWT. Mensagens com 300 bytes.

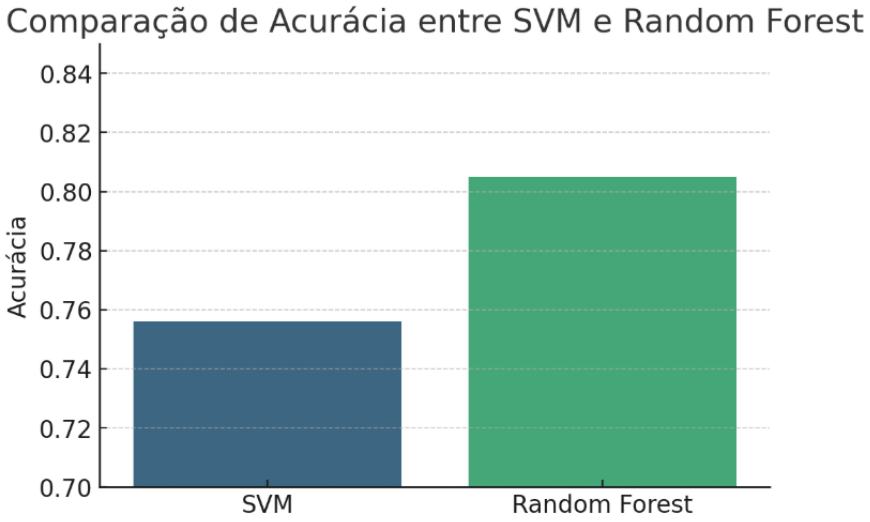

Figura 15: Acurácia SVM e Random Forest para DWT. Mensagens com 300 bytes.

apresenta características mais evidentes para análise forense, os métodos DCT e DWT tendem a ser mais "discretos", dificultando a tarefa dos classificadores.

5.3 Discussão dos Resultados

A comparação entre os três métodos evidencia um trade-off entre simplicidade, robustez e imperceptibilidade:

LSB é simples e de fácil implementação, mas sua fragilidade diante de compressões e ataques forenses limita sua aplicação em ambientes onde a segurança dos dados é crítica.

DCT oferece uma boa resistência à compressão, sendo vantajosa para aplicações que envolvem imagens em formatos como JPEG, embora exija cuidado na análise dos coeficientes para evitar a detecção.

DWT se destaca pela sua capacidade de trabalhar em múltiplas escalas de frequência, proporcionando um equilíbrio entre imperceptibilidade e robustez, o que o torna promissor para aplicações que demandam alta confiabilidade na recuperação dos dados mesmo após edições.

Adicionalmente, os resultados dos classificadores sugerem que a detecção au-

tomática de imagens modificadas pode servir como ferramenta auxiliar em análises forenses digitais, principalmente para identificar possíveis usos maliciosos de esteganografia. No entanto, a eficácia desses métodos de detecção ainda depende do contexto e da natureza das alterações aplicadas.

6 Conclusão

É possível concluir que este trabalho foi de intensos desafios, mas muito recompensador, pois foi uma chance de ampliar os conhecimentos acerca de segurança da informação e sobre o que está em alta nos dias atuais. Além disso, foi satisfatório obter resultados e perceber que apesar de complexos, foram experimentos divertidos de serem analisados.

Portanto, a realização deste estudo consolidou o entendimento sobre esteganografia e suas aplicações, destacando a importância da pesquisa contínua na área. A análise dos métodos testados permitiu uma visão mais clara de suas vantagens e limitações, incentivando futuras investigações e aprimoramentos. Dessa forma, este trabalho contribui para o entendimento prático e reforça a esteganografia como uma técnica valiosa na proteção da informação digital.

Referências

[1] Esteganografia. **Universidade Federal do Rio de Janeiro**. Disponível em: https://www.gta.ufrj.br/grad/09_1/versao-final/stegano/esteganalise.html.

[2] LAM, Vinh . Criar e impedir a esteganografia em cinco minutos. **OPSWAT**, 2023. Disponível em: https://portugese.opswat.com/blog/create-and-prevent-steganography-in-five-minutes.

[3] O que é esteganografia? Definição e explicação. kapersky. Disponível em: https://www.kaspersky.com.br/resource-center/definitions/what-is-steganography.

[4] LSB: Least Significant Bits. **WikiSec**. Disponível em: https://wiki.imesec.ime.usp.br/books/ctf-starter-pack/page/lsb-least-significant-bits.

[5] Comparative Analysis between DCT & DWT Techniques of Image Compression. **Journal of Information Engineering and Applications**. Disponível em: https://core.ac.uk/download/pdf/234676913.pdf.