



UNIVERSIDADE FEDERAL DE PERNAMBUCO
CENTRO DE TECNOLOGIA E GEOCIÊNCIAS
DEPARTAMENTO DE ENGENHARIA DE PRODUÇÃO
PROGRAMA DE PÓS-GRADUAÇÃO EM ENGENHARIA DE PRODUÇÃO

CAIQUE EMANUEL DA SILVA NUNES

**DESENVOLVIMENTO DE MÉTODO BASEADO EM APRENDIZAGEM DE
MÁQUINA PARA DETECÇÃO DE ÓLEO E LOOK-ALIKES EM IMAGENS**

Recife

2024

CAIQUE EMANUEL DA SILVA NUNES

DESENVOLVIMENTO DE MÉTODO BASEADO EM APRENDIZAGEM DE MÁQUINA PARA DETECÇÃO DE ÓLEO E LOOK-ALIKES EM IMAGENS

Dissertação apresentada ao Programa de Pós-Graduação em Engenharia da Produção da Universidade Federal de Pernambuco, como requisito parcial para a obtenção do título de Mestre em Engenharia de Produção.

Área de concentração: Pesquisa Operacional.

Orientadora: Prof^a. Dr^a. Isis Didier Lins.

Recife

2024

Catálogo na fonte:

Bibliotecário Carlos Moura, CRB-4/1502

N972d

Nunes, Caique Emanuel da Silva.

Desenvolvimento de método baseado em aprendizagem de máquina para detecção de óleo e look-alikes em imagens. / Caique Emanuel da Silva Nunes. – 2024.

85 f.: il.

Orientador: Profa. Dra. Isis Didier Lins.

Dissertação (mestrado) – Universidade Federal de Pernambuco. CTG. Programa de Pós-Graduação em Engenharia de Produção. Recife, 2024.

Inclui referências e apêndices.

1. Engenharia de produção. 2. Vazamento de óleo. 3. Look-alike. 4. Máquinas de vetores de suporte. 5. Redes Neurais convolucionais. 6. Classificação multi-rótulo. 7. Autoencoder variacional. I. Lins, Isis Didier (orientadora). II. Título.

CAIQUE EMANUEL DA SILVA NUNES

DESENVOLVIMENTO DE MÉTODO BASEADO EM APRENDIZAGEM DE MÁQUINA PARA DETECÇÃO DE ÓLEO E LOOK-ALIKES EM IMAGENS

Dissertação apresentada ao Programa de Pós-Graduação em Engenharia de Produção da Universidade Federal de Pernambuco, Centro de Tecnologia e Geociências, como requisito parcial para a obtenção do título de Mestre em Engenharia de Produção. Área de concentração: Pesquisa Operacional.

Aprovada em: 26 / 02 / 2024.

BANCA EXAMINADORA

Prof^ª. Dr^ª. Isis Didier Lins (Orientadora)
Universidade Federal de Pernambuco

Prof. Dr. Márcio José das Chagas Moura (Examinador Interno)
Universidade Federal de Pernambuco

Prof. Dr. Luis Felipe Ferreira de Mendonça (Examinador Externo)
Universidade Federal da Bahia

À minha mãe Selma e ao meu pai Manoel.

AGRADECIMENTOS

Quero agradecer primeiramente a Força maior que permite que as coisas aconteçam numa simples perfeição.

Agradeço aos meus pais, Selma e Manoel, por todo esforço, dedicação e amor incondicional que permitiu que eu chegasse aonde estou e por, juntamente com as minhas irmãs Dara e Midrache, acreditarem no meu potencial, muitas vezes até mais do que eu. Não poderia escolher pessoas melhores para compartilhar a minha vida.

À minha orientadora, a professora DSc. Isis Didier Lins, pela parceria, sabedoria e compartilhamento de experiências enriquecedoras. Estendo este agradecimento ao professor Márcio Moura e ao corpo docente do Programa de Pós-Graduação em Engenharia de Produção da Universidade Federal de Pernambuco, os quais admiro bastante.

Gostaria de agradecer também ao professor Paulo Estevão, por seu apoio e entusiasmo durante todo o mestrado, e a professora Maisa Mendonça, que me acompanha desde a época da graduação, por sua genialidade nas formas de enxergar as coisas da vida.

Aos meus amigos de turma João, Carlos, Danilo e Juliana pelo tanto que passamos juntos nesses últimos anos.

Agradeço também a todos que fazem parte do CEERMA, Lavínia, Thaís, July, Plínio e Leonardo pelo excelente time que formamos.

Ao Programa de Recursos Humanos (PRH 38.1) administrado pela Agência Brasileira do Petróleo, Gás Natural e Biocombustíveis (ANP) e pela Financiadora de Estudos e Projetos (FINEP); ao Conselho Nacional de Desenvolvimento Científico e Tecnológico (CNPq), e à Coordenação de Aperfeiçoamento de Pessoal de Nível Superior (CAPES).

A todos vocês que me inspiram a ser melhor a cada dia, meu muito obrigado!

RESUMO

Os oceanos, um dos principais ecossistemas do planeta, enfrentam diversas ameaças ao seu equilíbrio, sendo o vazamento de óleo um dos mais graves. Este estudo se propõe a utilizar a aprendizagem de máquina como ferramenta para uma rápida identificação desses vazamentos com o objetivo de minimizar os danos e custos associados. Três modelos são explorados com esta finalidade: Máquinas de Vetores de Suporte (SVM), um algoritmo muito eficaz em tarefas de classificação, Redes Neurais Convolucionais (CNN), que são altamente especializadas em tarefas de visão computacional, e Autoencoder Variacional (VAE), método reconhecido pela capacidade de detecção de anomalias. Os dois primeiros modelos foram construídos para realizar uma classificação multi-rótulo das imagens numa abordagem contextualizada, levando em consideração elementos como áreas costeiras, navios e look-alikes. Look-alikes representam um desafio ao objetivo final desta pesquisa, uma vez que se manifestam como manchas escuras nas imagens capturadas por satélites, de forma similar às manchas de vazamentos de óleo, contudo, sua origem não está relacionada a atividades humanas. A perspectiva da classificação multi-rótulo se apresenta como crucial na busca por tomar decisões direcionadas e evitar falsos positivos. A CNN destacou-se por sua capacidade de identificar múltiplas etiquetas multi-rótulo. Já o SVM tem a tendência de gerar proporções de acertos mais elevadas para cada etiqueta específica. E os resultados do VAE demonstram uma reconstrução altamente precisa. No final, espera-se que estes modelos possam trabalhar conjuntamente através de um sistema de votação para melhorar a detecção de vazamentos.

Palavras-chave: vazamento de óleo; look-alike; máquinas de vetores de suporte; redes neurais convolucionais; classificação multi-rótulo; autoencoder variacional.

ABSTRACT

The oceans, one of the planet's main ecosystems, face several threats to their balance, with oil spills being one of the most serious. This study proposes to use machine learning as a tool for quickly identifying these leaks with the aim of minimizing the damage and associated costs. Three models are explored for this purpose: Support Vector Machines (SVM), a very effective algorithm in classification tasks, Convolutional Neural Networks (CNN), which are highly specialized in computer vision tasks, and Variational Autoencoder (VAE), method recognized for its ability to detect anomalies. The first two models were built to perform a multi-label classification of images in a contextualized approach, taking into account elements such as coastal areas, ships and look-alikes. Look-alikes represent a challenge to the ultimate objective of this research, since they appear as dark spots in images captured by satellites, similar to oil spill spots, however, their origin is not related to human activities. The perspective of multi-label classification is crucial in the quest to make targeted decisions and avoid false positives. CNN stood out for its ability to identify multiple multi-label tags. SVM, on the other hand, tends to generate higher hit proportions for each specific tag. And the VAE results demonstrate a highly accurate reconstruction. In the end, it is expected that these models can work together through a voting system to improve leak detection.

Keywords: oil spill; look-alike; support vector machine; convolutional neural networking; multilabel classification, variational autoencoder.

LISTA DE FIGURAS

Figura 1 – Sistema Integrado	17
Figura 2 – Funcionamento SAR	23
Figura 3 – Classificador SVM.....	27
Figura 4 - Estratégia One-Vs-Rest	28
Figura 5 - Arquitetura básica da CNN para classificação multi-rótulo	30
Figura 6 – Camada de convolução.....	30
Figura 7 – Camada de pooling	31
Figura 8 – Padding	32
Figura 9 - Stride.....	32
Figura 10 - Autoencoder.....	34
Figura 11 - Sistema geral de detecção de vazamento	41
Figura 12 - Fluxograma da metodologia.....	43
Figura 13 – Imagem do Oil Spill Detection Dataset e sua máscara	46
Figura 14 - Arquitetura da CNN.....	52
Figura 15 – Arquitetura do VAE.....	53
Figura 16 - Sistema completo.....	63

LISTA DE TABELAS

Tabela 1 – Classes de look-alike que serão consideradas	47
Tabela 2 – Métricas para avaliação dos modelos	58
Tabela 3 - Proporção de acertos do modelo SVM.....	59
Tabela 4 – Proporção de acertos do modelo SVM.....	60
Tabela 5 – Proporção de acertos do modelo CNN.....	60
Tabela 6 - Métricas de avaliação do modelo VAE.....	61
Tabela 7 – Teste adicional VAE	62

LISTA DE SIGLAS

AE	Autoencoder
CNN	Redes Neurais Convolucionais
EMSA	European Maritime Safety Agency
ESG	Environmental, Social, and Corporate Governance
ML	Machine Learning
SAR	Synthetic Aperture Radar
SVM	Máquinas de Vetores de Suporte
O&G	Óleo e Gás
OTC	Offshore Technology Conference
OvA	One-Vs-All
OvR	One-Vs-Rest
VAE	Variational Autoencoder

SUMÁRIO

1	INTRODUÇÃO	13
1.1	OBJETIVOS	17
1.1.1	Objetivo Geral	17
1.1.2	Objetivos Específicos	18
1.2	JUSTIFICATIVA	18
1.3	ESTRUTURA DO TRABALHO	20
2	REFERENCIAL TEÓRICO E REVISÃO DA LITERATURA	22
2.1	SENSORIAMENTO REMOTO E LOOK-ALIKES	22
2.2	MODELOS DE APRENDIZAGEM DE MÁQUINA	24
2.2.1	Classificação de Imagens	24
2.2.2	SVM	27
2.2.3	CNN	29
2.2.4	VAE	33
2.2.5	Métricas de Avaliação	36
2.4	APLICAÇÕES DE APRENDIZAGEM DE MÁQUINA NA INDÚSTRIA DE ÓLEO E GÁS	38
3	METODOLOGIA	43
3.1	CONJUNTOS DE DADOS	45
3.1.1	Oil Spill Detection Dataset	45
3.1.2	TenGeoP – SARwv	46
3.1.3	Método 1 – utilizando SVM	48
3.1.4	Método 2 – utilizando CNN	50
3.1.5	Método 3 – utilizando VAE	53
4	ANÁLISE E DISCUSSÃO DOS RESULTADOS	58
5	CONCLUSÃO	64
5.1	LIMITAÇÕES	65

5.2	SUGESTÕES PARA TRABALHOS FUTUROS	66
	REFERÊNCIAS.....	68
	APÊNDICE A – MATRIZES DE CONFUSÃO MÉTODO 1 (APENAS BANCO DE DADOS PRINCIPAL)	75
	APÊNDICE B – MATRIZES DE CONFUSÃO MÉTODO 2 (APENAS BANCO DE DADOS PRINCIPAL)	77
	APÊNDICE C – MATRIZES DE CONFUSÃO MÉTODO 1 (BANCO DE DADOS PRINCIPAL + BANCO DE DADOS COMPLEMENTAR).....	80
	APÊNDICE D – MATRIZES DE CONFUSÃO MÉTODO 2 (BANCO DE DADOS PRINCIPAL + BANCO DE DADOS COMPLEMENTAR).....	83

1 INTRODUÇÃO

Cada vez mais, processos e experiências estão migrando para o ambiente digital, com o objetivo de simplificá-los e melhorar a qualidade dos resultados (VERHOEF et al., 2021). Impulsionada pela crescente capacidade de processamento das máquinas, a digitalização passou uma aceleração significativa nos últimos anos, estabelecendo fundamentos que podem ser considerados uma verdadeira revolução histórica.

Torna-se evidente o interesse tanto da indústria quanto da comunidade acadêmica na criação de tecnologias de inteligência artificial capazes de manejar com volumes massivos de dados e parâmetros simultaneamente, gerando resultados otimizados. Este campo de conhecimento oferece amplas oportunidades de aplicação (JORDAN; MITCHELL, 2015; LECUN; BENGIO; HINTON, 2015; SIDDIQUI; AMER, 2023) e inúmeros benefícios potenciais, o que ressalta a importância de seu domínio e desenvolvimento.

Outros fenômenos que vêm modificando a vida cotidiana são as mudanças climáticas. As variações drásticas de temperatura impactaram as percepções da sociedade, que passou a se planejar para lidar com este tipo de situação. Com a finalidade de evitar grandes consequências, é notória a crescente conscientização sobre a necessidade de alterações nas matrizes energéticas e nos padrões de consumo e produção.

Embora os possíveis desastres sejam provocados por processos inerentes à natureza, a gravidade de seus impactos pode ser intensificada por fatores de origem humana (como práticas de construção inadequadas, extração de recursos naturais e poluição), levando a uma pressão cada vez maior sobre a indústria e os governos para adotar medidas eficazes de prevenção, detecção e mitigação de riscos ambientais e sociais. Essa preocupação se estende por vários cenários e contextos dos quais vários ecossistemas são dependentes, como os oceanos, por exemplo.

Os oceanos enfrentam diversos riscos, sendo o derramamento de óleo um dos mais graves. As manchas de óleo nos oceanos podem surgir de processos conhecidos cientificamente como exsudação. Neste contexto, acontece uma liberação

do óleo do subsolo para a superfície marinha de forma não antropogênica devido à pressão subterrânea e à geologia local (JATIAULT et al., 2024).

Todavia, a principal causa de manchas de óleo nos oceanos são associadas a vazamentos que ocorrem em navios cargueiros ou nas plataformas offshore de petróleo (ANSELL et al., 2001; DE LA HUZ; LASTRA; LÓPEZ, 2019; DONG et al., 2022). O uso de barreiras de segurança é uma medida elementar de gerenciamento desse risco. Entretanto, em casos de falhas, é extremamente importante a rápida identificação e contenção do vazamento. Já que, à medida que o desastre se espalha, mais complicada e, conseqüentemente, mais cara é a sua mitigação.

Um suporte computacional se torna praticamente indispensável para tratar esta adversidade. Isso porque realizar de forma intensiva o monitoramento de oceanos, que possuem uma área de grande dimensão e representam um ecossistema complexo e dinâmico, com uma ampla gama de variáveis ambientais em constante modificação, seria quase impossível ou então uma operação muito dispendiosa se realizado sem assistência.

A identificação de vazamentos de óleo pode ser realizada a partir de imagens de satélite, por meio de análise e avaliação humana ou utilizando técnicas de visão computacional e inteligência artificial (VASCONCELOS et al., 2023). No entanto, a detecção de vazamentos torna-se difícil por conta da presença de look-alikes nas imagens (VASCONCELOS et al., 2020). Look-alikes são, segundo Topouzelis et al. (2007), regiões da imagem que têm estruturas de formações escuras com padrões parecidos com de vazamento de óleo. Contudo, as razões para as formações escuras, nesse caso, são naturais (como nuvens e áreas de frente de vento, por exemplo) e não ações humanas.

Nesse contexto, a aplicação de tecnologias de aprendizagem de máquina, especificamente de aprendizagem de máquina (ou machine learning – ML) representa uma solução acertada devido à habilidade de analisar e aprender a partir de grandes volumes de dados de maneira eficiente. Depois de treinado, o modelo de ML é capaz de reconhecer padrões e anomalias nas imagens, o que é crucial para diferenciar vazamento de óleo de semelhantes. O ML pode aprender com exemplos passados e, com o tempo, aprimorar a identificação de vazamentos com precisão, tornando-se uma ferramenta essencial para o gerenciamento de riscos ambientais.

Esse tipo de abordagem não apenas melhora a eficiência na detecção de vazamentos, mas também contribui para a redução do tempo de resposta em

situações críticas uma vez que a utilização de algoritmos de ML permite um monitoramento contínuo e automatizado de vastas extensões oceânicas. Isso se traduz numa identificação de incidentes envolvendo óleo em estágios iniciais, minimizando custos associados à mitigação de tais eventos.

É relevante destacar que as tecnologias de ML são versáteis e podem resolver uma ampla variedade de tarefas em diferentes domínios. A escolha de qual forma de resolver a tarefa será aplicada depende dos objetivos específicos do projeto e dos tipos de dados disponíveis.

Nesta dissertação, realiza-se a classificação multi-rótulo das imagens de radar de satélites Sentinel 1 geradas mediante uso da técnica Synthetic Aperture Radar (SAR). Isso significa que, em vez de simplesmente determinar se há um vazamento de óleo ou não e categorizar as imagens de forma binária, o sistema desta pesquisa será capaz de fornecer um conjunto de rótulos que descreverão os elementos detectados em cada imagem (LIZ et al., 2023).

Simultaneamente, várias características essenciais nas imagens poderão ser analisadas incluindo, a presença de vazamentos de óleo, de look-alikes, de áreas costeiras e de navios, além da delimitação da superfície do mar nas imagens. Em outras palavras, a classificação multi-rótulo leva em consideração o contexto da detecção.

A contextualização da detecção é crucial para entender as condições e o ambiente em que um vazamento de óleo ocorreu. Informações como a presença de áreas costeiras ou de navios, a título ilustrativo, são valiosas para tomar decisões mais direcionadas e assertivas sobre como responder ao vazamento e mitigar seus impactos.

Adicionalmente, ao incluir a capacidade de identificar look-alikes como uma categoria separada na classificação multi-rótulo, o sistema pode aprender a rotular essas áreas de forma mais apropriada. Isso é mais um fator que reduzirá a probabilidade de alarmes falsos e, conseqüentemente, evitará alocar recursos numa ação de resposta desnecessária.

Será explorada a aplicação de dois modelos distintos de ML para a classificação multi-rótulo das imagens nesta dissertação: Máquinas de Vetores de Suporte (SVM), representando uma abordagem mais tradicional de aprendizagem de máquina, e Redes Neurais Convolucionais (CNN), que utilizam técnicas de aprendizagem profunda.

O SVM é um algoritmo de aprendizado supervisionado que é muito eficaz em tarefas de classificação (CERVANTES et al., 2020b). Em sua forma mais básica, o SVM tenta encontrar um hiperplano de separação de maneira a maximizar a margem entre duas classes (UDDIN et al., 2019). Esta pesquisa busca estender o uso do classificador SVM para tarefas multi-rótulo, onde cada rótulo pode ser considerado como um problema de classificação binária separado (YAPP et al., 2020; ZHANG; YEUNG, 2013).

Já a CNN é uma classe de modelos de aprendizagem profundo altamente especializados em tarefas de visão computacional (SELVARAJU et al., 2020). Em sua essência, a CNN é projetada para extrair automaticamente características complexas de imagens (HU et al., 2020). Pretende-se avaliar o poder da CNN para detecção precisa e eficaz de múltiplos rótulos em imagens complexas.

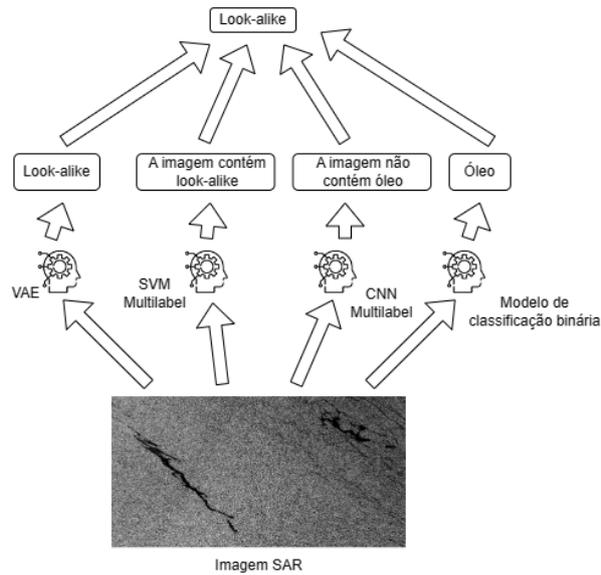
Essa dissertação também explorará a problemática sob uma perspectiva alternativa e apresentará também uma abordagem baseada em aprendizagem não supervisionada. Um modelo adicional foi desenvolvido empregando Variational Autoencoders (VAE).

O VAE tem a capacidade de aprender representações significativas dos dados e, a partir disso, gerar novas amostras semelhantes às observadas. Em outros termos, essa classe de modelos codifica reduzindo a dimensionalidade dos atributos e em seguida decodifica num processo inverso, associando essas operações a uma distribuição que regulariza o espaço latente de representação intermediária (MOUSAVIAN; EPPNER; FOX, 2019a).

A proposta consiste em treinar o modelo VAE usando exclusivamente imagens de look-alike, visto que eventos do tipo tendem a acontecer em maior quantidade. Uma vez treinado para compreender as nuances específicas dos look-alikes, o modelo terá uma tendência natural a gerar reconstruções menos precisas desses padrões quando comparadas a manchas de óleo.

É importante ressaltar que, no final, espera-se que tanto a classificação multi-rótulo quanto a abordagem não supervisionada possam desempenhar um papel agregador em um sistema abrangente, como mostra a figura 1. Neste sistema, a informação proveniente dos modelos SVM, CNN e VAE será sinergicamente considerada junto com informações derivadas de modelos de classificação binária tradicional.

Figura 1 – Sistema Integrado



Fonte: Autor (2024)

A ideia é combinar as previsões dos modelos de aprendizado de máquina para chegar a uma decisão final, por meio de um sistema de votação. A integração destes modelos visa aproveitar as vantagens de cada algoritmo, potencializando a precisão e abrangência na classificação das imagens de radar de satélites, e combiná-las num sistema de informação único. Esta seria uma forma de fortalecer a classificação automática e evitar falsos positivos. Dessa maneira, o sistema poderia se tornar mais sólido e eficiente na identificação de vazamento de óleo.

1.1 OBJETIVOS

1.1.1 Objetivo Geral

O objetivo geral deste trabalho é implementar uma classificação multi-rótulo, com base nas metodologias SVM e CNN, e uma abordagem não supervisionada, por meio de um modelo VAE, em imagens de monitoramento geradas por satélites Sentinel 1 que permitam, respectivamente, a identificação e diferenciação do derramamento de óleo e look-alikes.

1.1.2 Objetivos Específicos

O objetivo geral divide-se em outros, mais específicos:

- Ter acesso aos bancos de dados “Oil Spill Detection Dataset” (KRESTENITIS et al., 2019) e “TenGeoP – SARwv” (WANG et al., 2019), que são de domínio público, com imagens de monitoramento de satélites Sentinel 1 obtidas por meio da técnica Synthetic Aperture Radar (SAR);
- Aplicar modelos SVM e CNN que possam fazer uma classificação multi-rótulo assertiva e precisa das imagens de satélite;
- Implementar um modelo VAE capaz de aprender as características significativas das imagens de satélite, visando contribuir para a geração eficaz de amostras semelhantes às observadas;
- Realizar experimentos e testes de validação para avaliar a eficácia dos modelos propostos em termos de precisão;
- Analisar os resultados e destacar o benefício potencial de cada um dos modelos desenvolvidos neste trabalho para identificação de vazamento de óleo e look-alikes num sistema geral que opere em apoio a tomada de decisão em situações de desastres ambientais e na preservação de ecossistemas marinhos e costeiros.

1.2 JUSTIFICATIVA

O derramamento de óleo em mares e oceanos tem sido até hoje uma circunstância inevitável. De uma forma geral, eles não são planejados e quando acontecem provocam um desequilíbrio em todo o ecossistema envolvido. O vazamento de óleo afeta a biologia marinha e pode impactar negativamente a pesca e o turismo de uma região, à medida que se aproxima de uma área costeira. Um estudo que busque formas de minimizar estes danos se faz necessário por si só.

Além disso, é de se imaginar o quão custoso é um acontecimento indesejado na indústria de óleo e gás já que atividades rotineiras, como as de exploração e produção, costumam demandar investimentos sempre nas casas dos milhões de reais. Segundo a Agência Brasil, uma das empresas de comunicação estatal do país, os custos com a limpeza das praias e oceanos devido ao famoso acidente ocorrido

entre os anos de 2019 e 2020 no litoral brasileiro foram estimados em cerca de 188 milhões de reais (PLATONOW, 2021).

A tendência é que o vazamento de óleo no mar se torne cada vez mais oneroso, à medida que a responsabilidade das empresas do setor de petróleo em relação a esses incidentes aumenta. A cultura e consciência sustentável de governança ambiental, social e corporativa (do inglês, Environmental, Social, and Corporate Governance – ESG) está num patamar mais alto no ranking de investidores e governos do que apenas os ganhos financeiros de uma empresa.

O mundo caminha para o capitalismo de stakeholders e seguir a conduta ESG deve se tornar uma decisão estratégica para organizações que querem se manter atrativas (LITVINENKO et al., 2022; SHAKIL, 2021). Empresas que negligenciam os fatores social e de meio ambiente, ou seja que apresentam baixa pontuação ESG, já estão sendo penalizadas financeiramente (OOSTEROM; HALL, 2022).

Do ponto de vista monetário, também fica clara a extrema importância da rápida identificação e contenção do vazamento de óleo. Entretanto, não é uma tarefa fácil fazer o monitoramento de toda a extensão oceânica. Ainda mais porque as causas dos vazamentos de óleo no mar são diversas: vão desde condições corriqueiras de operação até situações de acidente graves em plataformas e navios de petróleo, o que acaba exigindo um acompanhamento constante.

Esse tipo de inspeção é viabilizada por meio da aplicação de inteligência artificial. Quando treinados, os modelos são capazes de receber imagens de satélite da área de interesse, processar essa informação e apontar se há ou não derramamento de óleo. A utilização de modelos de machine learning (ML) na digitalização de processos é uma área em progresso constante (ALZUBAIDI et al., 2021b).

Esta área de conhecimento tem tido muito interesse da academia e de indústrias por facilitar e tornar os processos mais assertivos. Realizar um estudo sobre o que tem sido feito até então no campo de visão computacional e aprendizagem de máquina para a identificação de óleo através de imagens, em como tornar os modelos mais eficientes e em como integrar as informações dos modelos é mais uma razão significativa para este trabalho.

Um diferencial deste trabalho é a atenção que será dada em definir uma maneira de distinguir o derramamento de óleo de look-alikes. O foco neste aspecto será um modo de evitar falsos diagnósticos. Após revisão de literatura, este se

mostrou ser um problema recorrente ainda pouco tratado, caracterizando uma oportunidade de pesquisa.

Outro diferencial do projeto será o desenvolvimento de algoritmos de classificação multi-rótulo. A aplicação de duas abordagens de classificação multi-rótulo permitirá uma análise mais detalhada e específica sobre o conteúdo das imagens. Ao considerar a presença ou ausência de múltiplos rótulos, em vez de apenas um, o sistema oferece uma resposta mais contextualizada e direcionada a diferentes aspectos da situação. Isso pode ser crucial para coordenar as ações em resposta a acidentes ambientais.

A incorporação de um modelo generativo baseado em Autoencoder Variacional (VAE) também oferecerá uma contribuição substancial na detecção automática de vazamento de óleo. Ao aprender uma representação latente das características distintivas dos look-alikes, o modelo demonstrará uma menor eficácia em reconstruir padrões que correspondam a manchas de óleo. Esta abordagem visa fortalecer a robustez do sistema abrangente por meio de uma diferenciação mais clara entre derramamentos reais e elementos semelhantes.

Dessa forma, esperam-se resultados mais aprimorados, com menos falsos diagnósticos, que proporcionem a diminuição de impactos ambientais causados pelos vazamentos de óleo e, conseqüentemente, a minimização de custos com reparos e melhora na tomada de decisão de ações.

1.3 ESTRUTURA DO TRABALHO

Além de um capítulo introdutório, que aborda a importância do uso de técnicas de visão computacional na identificação e diferenciação de vazamento de óleo e look-alikes, esta dissertação abrange outros quatro capítulos estruturados conforme descrito a seguir:

- **Capítulo 2:** apresenta a fundamentação teórica deste trabalho com conceitos de look-like, Suport Vector Machine (SVM), Redes Neurais Convolucionais (CNN), Variational Autoencoder (VAE), e de classificação multi-rótulo, essenciais para entendimento do estudo. Além disso, este capítulo fornece uma revisão da literatura, dando ênfase a como o conhecimento em estudo tem sido utilizado em pesquisas recentes.

- **Capítulo 3:** a metodologia desta pesquisa é detalhada neste capítulo, assim como os bancos de dados que são aplicados e os modelos criados para classificação e análises das imagens.
- **Capítulo 4:** exhibe e discute as métricas de avaliação de cada um dos modelos propostos e ainda analisa e compara os resultados retornados através de gráficos e tabelas.
- **Capítulo 5:** as principais contribuições da pesquisa e perspectivas conclusivas, e de forma semelhante comentários sobre as limitações e sugestões de trabalhos futuros, são expostas neste capítulo.

2 REFERENCIAL TEÓRICO E REVISÃO DA LITERATURA

Este capítulo irá abordar os conceitos fundamentais sobre os look-alikes, bem como explicar detalhadamente os modelos de Máquinas de Vetores de Suporte (SVM), Redes Neurais Convolucionais (CNN) e Variational Autoencoder (VAE). Também serão discutidas as métricas que serão empregadas na avaliação dos modelos propostos. Ao final deste capítulo, será apresentada uma seção de revisão da literatura na qual será explorado como esses conceitos têm sido aplicados na academia, demonstrando a relevância dessas abordagens em cenários práticos.

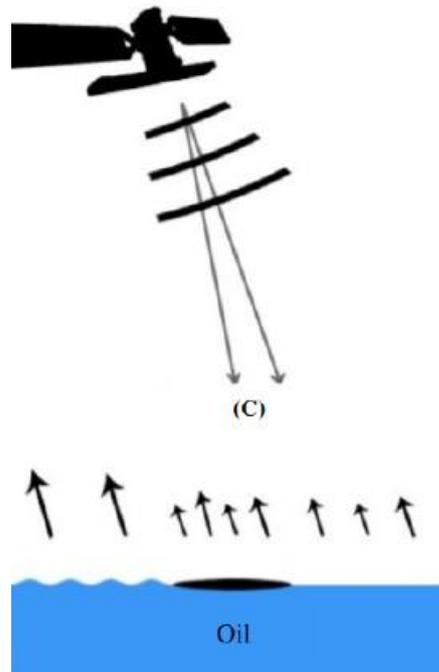
2.1 SENSORIAMENTO REMOTO E LOOK-ALIKES

O sensoriamento remoto é um procedimento que permite a coleta de dados sem a necessidade de contato direto com a superfície terrestre a partir de sensores localizados em plataformas distantes, como satélites, aeronaves ou drones. Existem diversos tipos de sensores remotos, cada um projetado para capturar informações específicas sobre o ambiente.

A principal tecnologia de sensoriamento remoto empregada para discernir alvos com densidades distintas, como no caso de óleo nos oceanos, é a geração de imagens de radar por meio da técnica Synthetic Aperture Radar (SAR). Devido as características desta técnica, imagens são produzidas independentemente do tempo cronológico, do clima e da iluminação (RAEISI; AKBARIZADEH; MAHMOUDI, 2018; YEKEEN; BALOGUN, 2020).

A Figura 2 mostra o funcionamento desta técnica e ajuda a entender a razão para a sua ampla aplicação para detecção de vazamento de óleo nos oceanos. O radar emite micro-ondas de energia que são retroespeelháveis. O óleo é um fluido de viscosidade maior do que a água, sendo, portanto, capaz de alterar a tensão superficial do oceano e de reduzir gradualmente a intensidade de ondas gravitacionais e capilares na área. Desta forma, em regiões onde há derramamento de óleo, o sinal retroespeelhado ao radar é menor do que em outras áreas da superfície do mar e por isso áreas escuras aparecem nas imagens geradas (AMRI et al., 2022; YEKEEN; BALOGUN, 2020).

Figura 2 – Funcionamento SAR



Fonte: (AMRI et al., 2022)

Entretanto, o vazamento de óleo não é a única causa de pontos escuros nas imagens. Nuvens de chuva, frentes de vento, correntes, gelo marinho e até mesmo ondas, entre outros fatores (ABOU SAMRA; ALI, 2022; GUO; ZHANG, 2014) também podem dar origem a manchas escuras em imagens SAR. Ainda assim, imagens SAR são amplamente utilizadas para este tipo de monitoramento porque são capazes de gerar imagens a uma longa distância sem qualquer imposição de hora nem de condições meteorológicas.

Essas áreas escuras nas imagens de monitoramento, semelhantes aos de vazamento de óleo, mas que tem uma origem natural são identificados como look-alikes. E, apesar de a etapa crucial num processo de identificação automático de vazamento de óleo ser distingui-lo de look-alikes (DHAVALIKAR; CHOUDHARI, 2021), é possível fazer a diferenciação dos pontos escuros a partir de algumas características.

Vazamentos de óleo e look-alikes apresentam propriedades físicas distintas e, portanto, reagem de formas diferentes sob certas condições (GUO; ZHANG, 2014). Liu et al. (2023) e Solberg et al. (1999) incorporam dados de algumas condições externas como informações a priori ao algoritmo de identificação.

As próprias características das áreas escuras podem contribuir para o discernimento do derramamento de óleo de look-alikes. Bhangale et al. (2017), Raeisi, Akbarizadeh e Mahmoudi (2018) e Yekeen e Balogun (2020) destacam a forma geométrica da área escura na imagem SAR como um importante atributo para extração de características. Isso se deve à facilidade de extrair características, mesmo em imagens de baixa qualidade, se comparado a outros atributos e a eficiência computacional em processá-las.

Valores médios, de desvio padrão, valores na escala de cinza, a complexidade, a solidez e a orientação são exemplos de características que podem ser extraídas da forma geométrica dos pontos escuros nas imagens SAR e que ajudam a observar melhor o fenômeno. A textura também é um exemplo de outros atributos que podem ser utilizados para extração de características no contexto em estudo.

2.2 MODELOS DE APRENDIZAGEM DE MÁQUINA

Os algoritmos de aprendizagem de máquina, assim como os de aprendizagem profunda, são ferramentas poderosas para resolver problemas complexos que envolvem uma grande quantidade de dados. Eles podem aprender padrões e relações a partir de exemplos e aplicá-los a novas situações. Um dos domínios mais desafiadores e interessantes para o uso de algoritmos de aprendizagem de máquina é a análise de imagens. Para extrair informações de imagens, os algoritmos podem realizar tarefas tais como: classificação, clusterização, recomendação, detecção de objetos e segmentação (MA et al., 2019; WU et al., 2021).

2.2.1 Classificação de Imagens

O termo classificação refere-se à ação de atribuir o rótulo apropriado a cada instância de dados (CHICCO; JURMAN, 2020). Os métodos de classificação variam a depender de como se pretende resolver o problema em questão. Num cenário no qual o objetivo é dividir as instâncias de dados em duas classes distintas usa-se uma classificação binária. A classificação binária é a base para os outros tipos de classificação.

Quando o número de possíveis classes para classificação é maior do que duas e deseja-se que a imagem seja classificada como apenas uma destas classes, emprega-se uma multiclassificação ou classificação multiclasse. O princípio dos algoritmos de multiclassificação é muito similar ao da classificação binária (YANG et al., 2014); a imagem será associada a classe que apresentar maior probabilidade.

Já quando o problema lida com situações em que várias classes simultaneamente podem ser atribuídas a cada instância de dados (como no caso de um dos objetivos específicos desta dissertação), trabalha-se com a classificação multi-rótulo, ou multilabel em inglês. No contexto de imagens, isso significa que uma imagem pode conter vários elementos que podem ser classificados em diferentes classes, e desta forma uma única imagem pode estar associada a várias classes simultaneamente.

A classificação multi-rótulo é aplicada em uma variedade de campos, incluindo processamento de linguagem natural (SHI et al., 2021), análise de texto (DUAN et al., 2022; QIU et al., 2020), reconhecimento de padrões em imagens (YONG LUO et al., 2013) e até mesmo em tarefas de classificação de produtos e recomendação em sistemas de comércio eletrônico (DENIZ; ERBAY; COŞAR, 2022). Em muitos desses casos, as instâncias podem ter várias características ou atributos que são importantes para descrever completamente o conteúdo ou o contexto.

Lidar com a classificação multilabel requer abordagens de aprendizagem de máquina especiais. Algoritmos tradicionais de classificação binária ou multiclassificação precisam ser adaptados ou estendidos para lidar com a previsão de várias etiquetas simultaneamente. Estratégias como a decomposição por conjuntos de pares, aprendizado hierárquico, uso de modelos pré-treinados e seleção de características relevantes são algumas das técnicas que podem ser empregadas para enfrentar os desafios da classificação multilabel (FÜRNKRANZ et al., 2008; ROUSU et al., 2006; SUN et al., 2021a, 2021b).

Para treinar um modelo de aprendizado de máquina em um problema multi-rótulo, é essencial fornecer ao modelo informações sobre quais classes estão presentes em cada amostra de treinamento, ou seja, imagem específica, a partir de um vetor binário. Nesse vetor, o valor indica que a classe está presente e o valor 0, que a classe está ausente. Matematicamente, seja $V = [v_1, v_2, \dots, v_n]$ um vetor de presença/ausência, em que n é o número de classes ou rótulos possíveis no problema multi-rótulo e $v_i \in \{0,1\}, \forall i$. Essa representação matemática do vetor de

presença/ausência permite ao modelo multilabel entender as classes associadas a cada amostra.

Diversos modelos de aprendizagem de máquina têm sido utilizados para fazer a classificação de imagens nos últimos anos. Particularmente, alguns desses modelos têm sido associados e contribuído com os avanços nas técnicas de sensoriamento remoto.

Estudos nas últimas duas décadas comprovam a eficácia do SVM em aplicações de sensoriamento remoto, o que o torna uma escolha sólida para aplicação nesta pesquisa. Alguns destes estudos ainda evidenciam a capacidade do SVM de lidar com imagens limitadas (MOSAVI et al., 2019; MOUNTRAKIS; IM; OGOLE, 2011; POKHARIYA; SINGH; PRAKASH, 2023; SHEYKHMUSA et al., 2020a).

Uma abordagem comum para adaptar SVMs para tarefas multilabel é usar a estratégia One-vs-Rest, que é uma alternativa ao método de decomposição por conjunto de pares. Neste caso, você treina um classificador SVM para cada classe, considerando-a como a classe positiva e todas as outras como a classe negativa. As previsões de cada classificador são então combinadas para obter as previsões finais.

Entretanto, a limitação de uso de outros modelos foi eliminada com avanços na tecnologia que permitiram o aumento na resolução das imagens de radar de satélite. A CNN, por exemplo, tem se mostrado um modelo ideal para lidar com a complexidade das imagens de alta resolução de radar de satélite pois são conhecidas por sua habilidade de classificação de elementos complexos em imagens (CHENG; ZHOU; HAN, 2016; HU et al., 2015).

Modelos CNN incorporam de forma inerente estratégias apropriadas para a classificação multilabel, as quais costumam estar implícitas na própria arquitetura da rede neural, sendo o aprendizado hierárquico uma destas. Na aprendizagem hierárquica, o problema de classificação é decomposto em níveis hierárquicos, onde cada nível representa uma camada da hierarquia. Os classificadores são treinados em cada nível para prever a presença ou ausência de rótulos nas classes associadas a esse nível. Essa abordagem pode ajudar a lidar com o problema de escalabilidade em conjuntos de dados com muitas classes, uma vez que o treinamento e a previsão podem ser realizados em níveis mais específicos da hierarquia, tornando o problema mais gerenciável (ROY; PANDA; ROY, 2020; XU; ZHU; TAPPER, 2020).

Inspirados pelos avanços em aprendizado profundo, os modelos VAE tem sido cada vez mais aplicados a problemas de visão computacional. Além disso, o VAE

também oferece uma solução dada a dificuldade em obter amostras rotuladas, como no caso dos tipos de look-alike, explorando distribuições latentes e gerando novas amostras sem a necessidade de grandes conjuntos de dados rotulados (KHALID; WOO, 2020; WANG et al., 2020a, 2020b; XIAN et al., 2019).

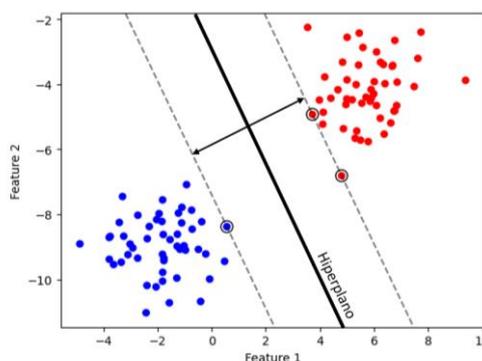
A combinação da aplicação de métodos baseados nestes três modelos (SVM, CNN e VAE) em um sistema abrangente se mostra uma abordagem disruptiva e inovativa, podendo otimizar o resultado final utilizando o melhor de cada modelo e obtendo respostas consideradas sinergicamente.

2.2.2 SVM

O Support Vector Machines - SVM são uma classe de algoritmos de aprendizagem de máquina que se destacam na classificação e regressão (CERVANTES et al., 2020b). Introduzido na década de 1960 e alcançando popularidade nos anos 1990 (KECMAN, 2005), o SVM ganhou popularidade devido à sua eficácia em lidar com problemas de classificação complexos, especialmente em conjuntos de dados de alta dimensão.

A figura a seguir apresenta o conceito central por trás das SVMs é criar um hiperplano que melhor separa as diferentes classes de dados no espaço de características (PISNER; SCHNYER, 2020). No caso mais simples, onde temos duas classes, o objetivo é encontrar um hiperplano que maximize a margem, ou seja, a distância entre o hiperplano e os exemplos mais próximos de cada classe. Esses exemplos são chamados de vetores de suporte, daí o nome "Máquinas de Vetores de Suporte".

Figura 3 – Classificador SVM



Fonte: (Adaptado de SIDEY-GIBBONS; SIDEY-GIBBONS, 2019)

A importância da margem é que ela ajuda a garantir uma boa generalização do modelo para novos dados, reduzindo o risco de overfitting. Além disso, SVMs podem lidar com conjuntos de dados não linearmente separáveis através da transformação dos dados em um espaço de maior dimensão onde eles se tornam linearmente separáveis. Isso é feito através de funções de kernel, que permitem calcular as relações de similaridade entre os pontos de dados nesse novo espaço (KAYA, 2013; POLAT; GÜNEŞ, 2009).

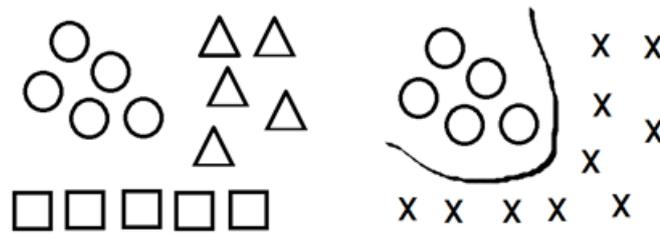
As SVMs também são frequentemente usadas para resolução de problemas de regressão, onde o objetivo é prever um valor contínuo em vez de uma classe. Nesse caso, o hiperplano é ajustado de forma a minimizar o erro entre as previsões do modelo e os valores reais dos dados de treinamento (CERVANTES et al., 2020a; LIANG et al., 2013).

Apesar de sua eficácia, as SVMs têm algumas desvantagens. Elas podem se tornar computacionalmente caras em conjuntos de dados muito grandes, e a escolha do kernel apropriado pode ser desafiadora. Além disso, a interpretabilidade dos resultados de uma SVM pode ser complexa, especialmente quando a dimensionalidade dos dados aumenta (CERVANTES et al., 2020a; SHEYKHMOUSA et al., 2020b).

No geral, as Máquinas de Vetores de Suporte são uma ferramenta poderosa no arsenal tradicional de técnicas de aprendizado de máquina, com aplicação em uma ampla gama de problemas. Sua capacidade de lidar com dados complexos e não lineares, bem como sua ênfase na maximização da margem, contribuem para sua utilidade em projetos que envolvem classificação, regressão e até mesmo detecção de anomalias (CHANG et al., 2020; GHODDUSI; CREAMER; RAFIZADEH, 2019; SHEYKHMOUSA et al., 2020b).

A adaptação do SVM binário para o cenário multi-rótulo envolve uma abordagem chamada "One-Vs-Rest" (OvR) ou "One-Vs-All" (OvA) apresentada na figura 4. É importante destacar que, como cada classe é tratada individualmente em relação a todas as outras classes, não em relação a pares específicos, esta não é uma técnica de decomposição por pares, mas sim uma adaptação direta de classificadores SVM binários para lidar com múltiplos rótulos.

Figura 4 - Estratégia One-Vs-Rest



Fonte: (Adaptado de YANG et al., 2014)

Nesse contexto, o objetivo é treinar um classificador SVM binário para cada classe individualmente, considerando-a como uma classe positiva, enquanto todas as outras classes são tratadas como uma classe negativa. Esse processo é repetido para cada classe presente no problema multi-rótulo.

2.2.3 CNN

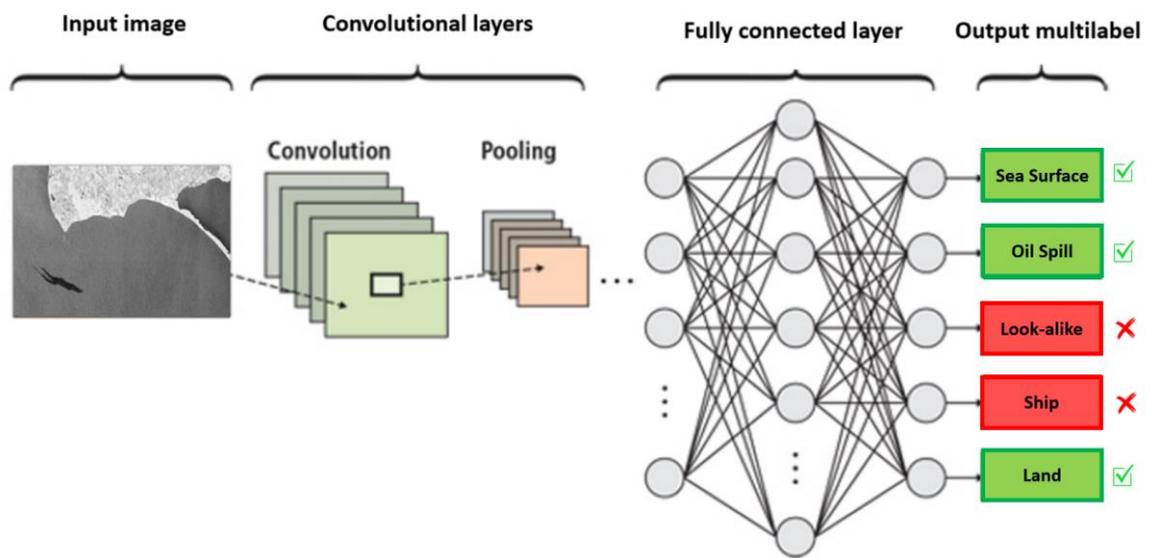
Em contraste com o modelo tradicional de aprendizagem de máquina mencionados previamente, esta seção se volta para a exploração das Redes Neurais Convolucionais, ou Convolutional Neural Network (CNN). As CNNs são uma classe especializada de redes neurais profundas amplamente utilizadas em tarefas de visão computacional e processamento de imagem. Modelos deste tipo de rede neural têm se destacado ao automatizar o processo de extração de características e aprender hierarquicamente representações cada vez mais abstratas, tornando-as especialmente adequadas para tarefas de classificação multi-rótulo em imagens (AL BATAINEH et al., 2023; BHATT et al., 2021; DONG et al., 2014).

As CNNs foram introduzidas por Yann LeCun et al. em 1989 (LECUN et al., 1989; LECUN; KAVUKCUOGLU; FARABET, 2010), mas ganharam destaque significativo nas últimas décadas, devido ao seu desempenho em uma ampla gama de aplicações, como classificação de imagens, detecção de objetos e segmentação de imagens. E apesar dos dados de input geralmente serem imagens, também pode-se trabalhar com vetores de características texto e sons neste tipo de modelo.

A arquitetura das CNNs é inspirada na organização do córtex visual do cérebro humano e tem o objetivo de representar em máquinas a forma como reconhecemos imagens. Elas consistem em várias camadas, cada uma desempenhando um papel

específico no processo de extração e aprendizado de características das imagens. As principais camadas em uma CNN incluem: a camada de convolução, de pooling e totalmente conectada. A figura a seguir ilustra a arquitetura básica de uma CNN, destacando a interconexão das camadas na extração e aprendizado das características das imagens (ALZUBAIDI et al., 2021a; ZHAO et al., 2019).

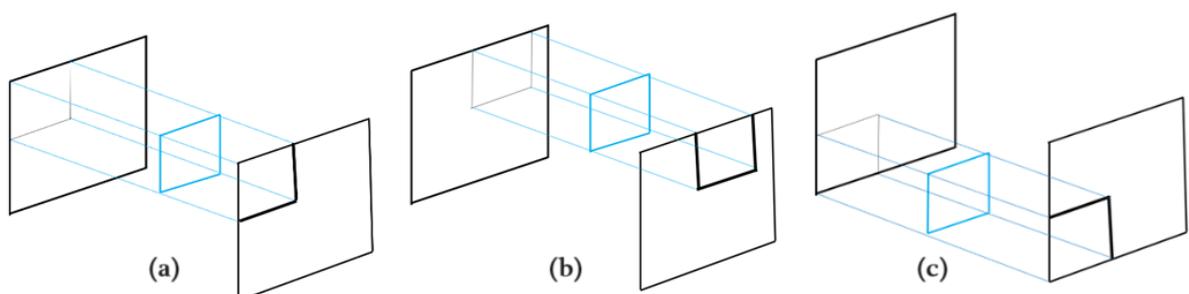
Figura 5 - Arquitetura básica da CNN para classificação multi-rótulo



Fonte: Autor (2024)

A camada de convolução é o bloco fundamental de uma CNN. Ela opera aplicando filtros (kernels) a regiões locais da imagem de entrada. Esses filtros deslizam pela imagem e realizam operações de convolução para detectar padrões e características relevantes, como mostra a figura 6.

Figura 6 – Camada de convolução

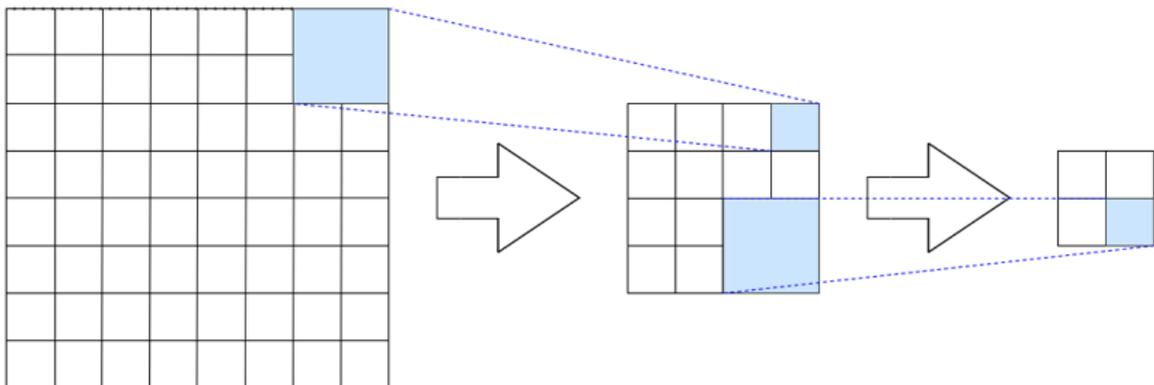


Fonte: Autor (2024)

Pode-se observar na figura 6 que o kernel, representando na cor azul, deslizando pela imagem em três etapas (a), (b), e (c), resultando na geração da convolução. Matematicamente, teremos uma convolução gerada toda vez que o filtro passa por uma pequena quantidade de pixels e é dessa forma que um padrão desejado é identificado. Uma convolução é uma operação matemática em que dois objetos produzem um output que mostra como a forma de um é modificado pelo outro (GOODFELLOW; BENGIO; COURVILLE, 2016). A saída dessa camada é chamada de mapa de características ou feature map.

A camada de pooling (ou subamostragem), que é apresentada na figura 7, é usada para reduzir a dimensionalidade do mapa de características, eliminando ruídos e preservando as características mais importantes. A técnica mais comum é o max-pooling, que seleciona o valor máximo em cada região de pooling (ALZUBAIDI et al., 2021a; ZHAO et al., 2019).

Figura 7 – Camada de pooling



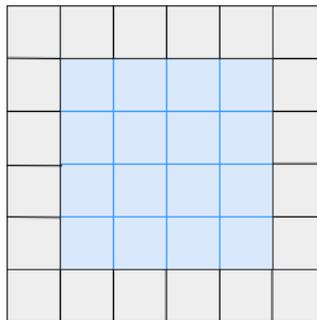
Fonte: Autor (2024)

Como mostra a figura 7, a camada de pooling consiste na aplicação de uma operação de agregação que varia a depender da técnica escolhida, podendo ser média ou máximo por exemplo. Em outras palavras, uma região de entrada é resumida a um valor único que melhor a representa.

Após a extração de características por meio de camadas de convolução e pooling, as camadas totalmente conectadas são usadas para realizar a classificação final. Essas camadas são semelhantes às encontradas em redes neurais tradicionais e combinam as características extraídas para produzir uma saída final.

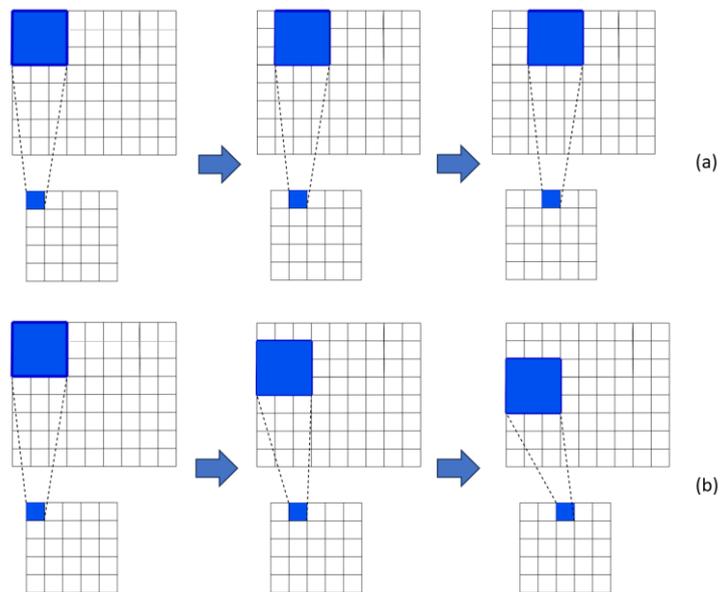
Existem alguns mecanismos que auxiliam nos ajustes do modelo CNN. O padding, figura 8, por exemplo, adiciona pixels nas bordas dos dados para que a informação desta localização não seja desvalorizada. Já o stride, figura 9, define a forma como o filtro irá se movimentar entre os dados (horizontal e verticalmente) (ALZUBAIDI et al., 2021a; ZHAO et al., 2019).

Figura 8 – Padding



Fonte: Autor (2024)

Figura 9 - Stride



Fonte: Autor (2024)

Os dois conceitos dos mecanismos abordados da CNN estão ilustrados nas imagens anteriores. Na figura 8, uma camada representada em tons de cinza é anexada a imagem de input em azul. A extensão da imagem se dá como forma de evitar a perda de informação de quando o filtro se move em direção as bordas da

imagem e encontra menos pixels para serem combinados. O padding serve, por conseguinte, para que a informação das bordas da imagem seja considerada para convolução da mesma forma que a informação do centro da imagem.

A figura 9 expõe a sequência em que o filtro se move ao longo da imagem. Mais especificamente, na figura 9.a o filtro se desloca horizontalmente enquanto que na figura 9.b o movimento é vertical. A forma e, além disso, a quantidade de pixels que serão deslocados em cada etapa do percorrer do filtro é determinada com o uso do stride.

O treinamento de uma CNN envolve a otimização dos pesos dos neurônios para minimizar uma função de perda, geralmente usando algoritmos de otimização. Desta forma, durante o treinamento, as CNNs aprendem a extrair automaticamente as características mais discriminativas das imagens para realizar tarefas específicas, como classificação.

No final, a saída de uma CNN é uma distribuição de probabilidade sobre as classes-alvos. Em outras palavras, a saída consistirá em um vetor de pontuações, onde cada elemento do vetor corresponde a uma classe possível. Essas pontuações representam o grau de confiança da CNN de que a imagem pertence a cada classe.

Posteriormente, é comum aplicar uma função softmax à saída da CNN para converter essas pontuações em probabilidades. A classe com a maior probabilidade é considerada a predição final da CNN. Na classificação binária, a resposta é um único valor que indica a probabilidade ou confiança de uma única classe. Em contraste, na classificação multi-rótulo, a resposta é um vetor binário que reflete a presença ou ausência de várias classes em uma única imagem.

As redes neurais convolucionais e suas variantes têm sido amplamente aplicadas em diversas áreas, incluindo diagnóstico médico por meio da análise de imagens de exames (MAIOR et al., 2021; SHARMA; PARASHAR; JOSHI, 2021) e no reconhecimento de objetos na construção de veículos autônomos (LI et al., 2020).

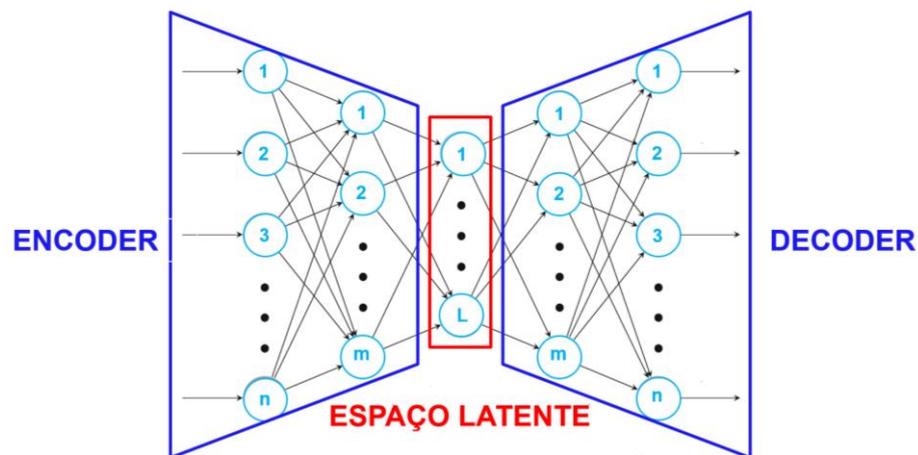
2.2.4 VAE

Muitas vezes, os conjuntos de dados disponíveis para estudo não possuem rótulos adequados (LIU et al., 2017). Há, entretanto, um ramo do aprendizado de máquina que permite encontrar padrões e relações úteis nos dados sem usar rótulos ou supervisão prévia conhecido como aprendizagem não supervisionada.

A aprendizagem não supervisionada oferece, na maioria das vezes, uma maneira de extrair informações ao simplificar a representação dos dados a partir da redução de dimensionalidade, que consiste em transformar os dados originais numa representação mais compacta e relevante, preservando as características essenciais.

O autoencoder (AE) é uma classe de modelos de aprendizado de máquina não supervisionado que demonstra uma ligação estreita com técnicas de redução de dimensionalidade. Trata-se de redes neurais artificiais que aprendem a codificar os dados de entrada em uma representação latente de menor dimensão e depois a decodificar essa representação para reconstruir os dados originais (CHENG et al., 2020), como mostra a figura 10.

Figura 10 - Autoencoder



Fonte: (Adaptado de LOTFOLLAHI et al., 2020)

Como pode ser visto na figura 10, os AEs operam numa arquitetura básica que consiste em duas partes principais: o codificador (encoder), responsável por comprimir os dados de entrada numa representação latente de menor dimensionalidade, e o decodificador (decoder), cuja função é reconstruir os dados a partir da representação latente (LIU et al., 2017).

Um outro fator importante no desenho de um AE representado na figura 10 é o tamanho do gargalo do modelo, ou seja, o número de nós na camada de espaço latente. Essa camada funciona como um ponto de compressão crítico, determinando a capacidade do modelo em aprender as representações mais significativas dos dados (GONG et al., 2019a). Se o gargalo for muito grande, o AE pode memorizar os

dados e não aprender características relevantes. Se o gargalo for muito pequeno, o AE pode perder informação e não conseguir reconstruir os dados com precisão.

Os AEs diferem de outros modelos de redes neurais artificiais também em termos de treinamento. O foco do AE é em minimizar o erro de reconstrução, garantindo que a saída do decodificador seja o mais próximo possível da entrada original (BENGIO; COURVILLE; VINCENT, 2013).

Esse critério de treinamento permite que o AE seja empregado em tarefas de classificação. Dado que o modelo é treinado para minimizar o erro de reconstrução, espera-se um erro maior para dados que gerem reconstruções que desviem do comportamento típico. Desta forma, pode-se, então, definir um limiar para classificar os dados como anomalias ou não. Além disso, a saída do codificador, considerada um conjunto de características discriminativas de alto nível, também pode ser empregada em tarefas de classificação (GONG et al., 2019b; LOTFOLLAHI et al., 2020).

Embora sejam ferramentas valiosas em aprendizado de máquina, os autoencoders tradicionais podem apresentar algumas limitações em relação à representação latente e geração de dados. Essas limitações abrangem a capacidade de capturar representações complexas e não lineares, sensibilidade a ruídos e outliers nos dados, dificuldades na geração de dados realistas e limitações na manipulação de dados não estruturados ou sequenciais.

Na literatura, variantes mais sofisticadas de autoencoders são propostas quando se faz necessário considerar alguma dessas limitações. O Autoencoder Variacional (Variational Autoencoder – VAE), por exemplo, surge como resposta à limitação dos autoencoders tradicionais em que a representação latente gerada por um modelo é caracterizada por uma dimensão fixa (BOWMAN et al., [s.d.]).

Essa limitação é particularmente relevante em situações em que os dados são altamente complexos e apresentam uma diversidade considerável. A variabilidade dos dados não pode ser completamente condensada em uma representação latente de dimensão fixa, pois isso pode resultar em uma perda de detalhes e nuances.

No VAE, em vez de produzir um único ponto no espaço latente, o encoder gera parâmetros que descrevem uma distribuição, frequentemente assumindo a forma de uma distribuição normal (gaussiana), para aprender uma representação eficiente e probabilística dos dados (BROCKI; CHUNG, 2019; ZEROUAL et al., 2020).

Em outras palavras, os VAEs são modelos generativos que contém técnicas probabilísticas integradas aos conceitos e estrutura básica de um autoencoder. Ou seja, os modelos VAE não apenas aprendem representações latentes, mas também modelam a distribuição dos dados na representação latente. Essa abordagem proporciona uma compreensão mais profunda e flexível dos dados, permitindo uma melhor captura da incerteza inerente às observações (MOUSAVIAN; EPPNER; FOX, 2019b; SINHA; EBRAHIMI; DARRELL, 2019).

A incerteza inerente às observações não é necessariamente devida à ausência de rótulos nos dados. Embora a falta de rótulos possa contribuir em algumas situações, há vários outros fatores que podem levar à incerteza pois dados reais muitas vezes exibem uma variabilidade natural. Modelar a incerteza permite que o modelo reconheça e capture essa variabilidade intrínseca.

Visto que modelos que geram representações latentes de dimensão fixa podem não ser capazes de capturar todas as variabilidades, levando a uma simplificação excessiva e, em alguns casos, a uma representação menos fiel dos dados originais, este trabalho propõe a construção e uso de um modelo baseado em VAE para diferenciação de look-alikes e vazamento de óleo.

Em tarefas onde a classificação ou reconhecimento de padrões são desafiadores devido à similaridade entre diferentes classes ou à ambiguidade nas características, como é o caso, a modelagem da incerteza oferecerá uma maneira mais realista de representar as previsões do modelo.

2.2.5 Métricas de Avaliação

Cinco métricas de avaliação foram escolhidas para análise dos modelos de aprendizagem de máquina que fazem a classificação multilabel neste trabalho: precisão, taxa de verdadeiros positivos, F1-score, perda de Hamming e acurácia. As três primeiras métricas listadas serão calculadas considerando cada classe, enquanto que as duas últimas são calculadas para cada amostra individualmente no conjunto de teste. Em seguida, a média de cada métrica é obtida e, na posse destes valores, pode-se avaliar a capacidade do modelo de fazer previsões assertivas e comparar as diferentes abordagens que estão sendo estudadas.

Com relação à precisão, o modelo calcula a média das precisões para cada classe usando a equação 1 e, em seguida, tira a média dessas médias (average

precision), usando as previsões feitas pelo modelo e as labels verdadeiras correspondentes do conjunto de teste. Para o cálculo desta métrica são considerados tanto os verdadeiros positivos quanto os falsos positivos, ou seja, ela informa a porcentagem de predições corretas.

$$Precisão = \frac{Verdadeiros\ Positivos}{(Verdadeiros\ Positivos+Falsos\ Positivos)} \quad (1)$$

Similarmente, o modelo calcula a taxa média de verdadeiros positivos das classes individualmente, como mostra a equação 2, e depois faz a média dessas médias (average recall). Todavia, essa métrica mede a capacidade de o modelo identificar corretamente todas as instâncias relevantes, o que significa dizer que ela considera tanto os verdadeiros positivos quanto os falsos negativos.

$$Recall = \frac{Verdadeiros\ Positivos}{(Verdadeiros\ Positivos+Falsos\ Negativos)} \quad (2)$$

A F1-score é uma métrica que leva em conta as outras duas mencionadas antes, como mostra a equação 3. Trata-se de uma medida que reflete a capacidade do modelo fazer previsões precisas e identificar corretamente todas as instâncias relevantes. E por se tratar de um modelo multilabel, o modelo também computa a média das médias dos F1-scores de cada classe.

$$F1-Score = \frac{(2 \times Precisão \times Recall)}{(Precisão+Recall)} \quad (3)$$

A perda de Hamming (Hamming Loss) mede a fração média de labels incorretamente classificados para todas as amostras. Em outras palavras, ela avalia tanto os rótulos que foram erroneamente adicionados quanto os que foram omitidos. A única métrica dentre o grupo escolhido para avaliação do modelo que quanto menor melhor. A perda de Hamming é representada na equação 4:

$$Perda\ de\ Hamming = \frac{1}{N} \sum_{i=1}^N \frac{1}{m} \sum_{j=1}^m (y_{ij} \oplus \hat{y}_{ij}) \quad (4)$$

onde, N é o número de amostras, m é o número de rótulos, y_{ij} é o rótulo verdadeiro e \hat{y}_{ij} é o rótulo previsto. O operador \oplus denota a diferença simétrica (XOR), que compara dois valores e retorna ou 1 se os valores forem diferentes, ou 0 se forem iguais.

Adicionalmente, foi definida uma função no modelo para computar o quão bem o modelo é capaz de prever corretamente todas as labels para uma amostra específica (subset accuracy), equação 5. Em outros termos, ainda que o modelo classifique quatro das possíveis classes de uma imagem corretamente e apenas uma das classes não o seja, a classificação é considerada parcialmente correta e é ignorada por essa métrica.

$$\text{Subset Accuracy} = \frac{\text{N}^\circ \text{ de Amostras com todos os R\^otulos Corretos}}{\text{Total de Previs\^oes}} \quad (5)$$

Já para avaliar o modelo VAE será utilizada outra métrica: o erro médio de reconstrução. Essa métrica oferece não só uma visão abrangente do desempenho do VAE, mas também é essencial para entender a capacidade de reprodução do modelo e, portanto, sua capacidade de capturar e representar de maneira eficaz a variabilidade presente nos dados.

O erro médio de reconstrução é uma métrica fundamental que quantifica a discrepância entre as imagens de entrada e suas reconstruções geradas pelo VAE. É calculado como a média das diferenças absolutas pixel a pixel entre as imagens originais (X) e as reconstruções (X'). Uma baixa pontuação nessa métrica indica que o modelo é capaz de reproduzir com precisão as características das imagens de entrada, sendo uma medida direta da qualidade das reconstruções.

$$\text{Erro Médio de Reconstrução} = \frac{1}{N} \sum_{i=1}^N |X_i - X'_i| \quad (6)$$

2.4 APLICAÇÕES DE APRENDIZAGEM DE MÁQUINA NA INDÚSTRIA DE ÓLEO E GÁS

No ramo de energias, no setor de óleo e gás (O&G) mais especificamente, alguns modelos de ML já são utilizados e ajudam a identificar perfis, padrões e anomalias que beneficiam todos os setores desta cadeia produtiva, tanto do downstream (parte da cadeia logística que envolve o transporte e a distribuição dos produtos até os locais de consumo), como do midstream (relacionado às atividades de refino e transformação do petróleo em produtos), e do upstream (caracterizado pelas atividades de busca, localização e perfuração de poços de petróleo) da indústria.

Um exemplo é o projeto da Petrobras apresentado na Offshore Technology Conference (OTC) Brasil em 2019 (MICROSOFT NEWS CENTER BRASIL, 2019) que utiliza a inteligência artificial da Microsoft e câmeras de alta resolução para proporcionar um ambiente industrial mais seguro. Os algoritmos do projeto englobam ML e são capazes de identificar ameaças através de imagens. O resultado ainda é transformado em dados estatísticos que permitem um treinamento específico dos colaboradores e reduz riscos em toda a plataforma.

De forma prática, caso o sistema identifique uma situação de posicionamento inseguro de uma carga, ou de obstrução de uma rota de fuga, ou de acesso a ambientes restritos por pessoas não autorizadas ou até mesmo o uso errôneo de um equipamento de proteção individual, ele emitirá um alerta. Isso só é possível porque o sistema foi treinado com uma quantidade suficiente de imagens (de conjuntura normal e de risco) e aprendeu a identificá-las. Cerca de 300 pessoas (MICROSOFT NEWS CENTER BRASIL, 2019) são beneficiadas pelo uso de ML apenas na plataforma em discussão. Além da segurança, o sistema também garante a continuidade operacional e, portanto, uma maior produtividade.

Outros usos de modelos de ML na indústria de O&G estão ligados à área da manutenção de equipamentos. Jin et al. (2014), com o objetivo de aprimorar o gerenciamento de riscos de gasodutos e minimizar os custos com possíveis vazamentos, aplicaram modelos de ML para detectar e localizar vazamentos nos dutos de transporte de gás natural. Foram utilizados para isso dados acústicos do sistema.

Soluções automáticas e modelos de aprendizagem de máquina também têm sido testados no contexto de detecção de vazamento de óleo. Ronci et al. (2020) treinaram um modelo de CNN com uma função adversária e trouxeram como diferencial um método de translação de imagens com a finalidade melhorar os resultados classificatórios. No contexto de derramamento de óleo, onde as imagens reais são escassas e difíceis de obter, a translação de imagens se mostrou uma técnica capaz de melhorar o desempenho de modelos de aprendizado de máquina.

Todavia, para estudos cuja problemática estende o foco para look-alikes como é o caso, a translação de imagens pode não ser tão necessária pois já há uma quantidade e uma diversidade suficiente de dados. Dois estudos exemplificam e contribuem significativamente com esta ideia.

O primeiro foi realizado por Wang et al. (2019) e se concentrou em reunir, para exploração de modelos de machine learning, uma extensa base de imagens SAR de fenômenos geofísicos que podem ser considerados look-like. O segundo foi conduzido por Krestenitis et al. (2019) que direcionaram esforços para construir uma rede convolucional profunda e disponibilizaram publicamente o banco de dados de imagens construído como forma de contribuir com o aperfeiçoamento de modelos de detecção de vazamento de óleo de forma automática.

Já Yekeen e Balogun (YEKEEN; BALOGUN, 2020) fazem uma revisão do que tem sido feito até então na detecção e controle de vazamento de óleo e exploram as potenciais soluções de digitalização dessas operações. Diante dos casos estudados, os autores destacam tanto a eficácia do SVM em realizar uma classificação robusta ao lidar com conjuntos de dados de alta dimensionalidade, quanto a notável precisão na interpretação de dados complexos por modelos de aprendizagem profunda, como CNN e autoencoders.

As vulnerabilidades dos métodos que são aplicados atualmente também são expostas pelos autores, o que ajuda a entender onde os modelos de aprendizagem de máquina podem fazer suas contribuições. A principal limitação apontada está relacionada a dificuldade em distinguir vazamentos reais de look-alikes, o que torna os modelos revisados específicos para particularidades.

A pesquisa realizada por Cheng et al. (CHENG et al., 2020) vai além e elenca os principais obstáculos da classificação de imagens obtidas por sensoriamento remoto baseados em modelos generativos adversariais, CNN e autoencoders. Se sobressai a observação de que a classificação de uma imagem por um único rótulo não condiz muito com a realidade uma vez que é alta a frequência de imagens que contém múltiplos objetos sendo, portanto, uma classificação multilabel mais adequada. Também é enfatizado que é dispendioso etiquetar todas as imagens, sendo a aprendizagem não supervisionada uma alternativa eficiente.

Huby, Alubady e Sagban (2022) desenvolveram e testaram uma nova arquitetura de aprendizagem de máquina ao mesmo tempo em que buscavam por um tipo de pré-processamento que reduzisse o tempo de processamento das imagens e, conseqüentemente, necessitasse de um menor empenho computacional.

A nova arquitetura proposta pelos autores aprimora recursos da CNN por meio de modelos generativos, que possuem como conceito base a redução e reconstrução dos dados. A melhoria nos resultados das métricas avaliativas enfatizam a força que a combinação entre modelos pode proporcionar.

A pesquisa realizada por Negreiros et al. (2022) examinou profundamente vários aspectos do vazamento de óleo nos oceanos, desde propriedades físicas e químicas até os tipos de detecção, manual e automática. Além disso, algumas lacunas foram apontadas pelos autores como potenciais linhas de aperfeiçoamento da área.

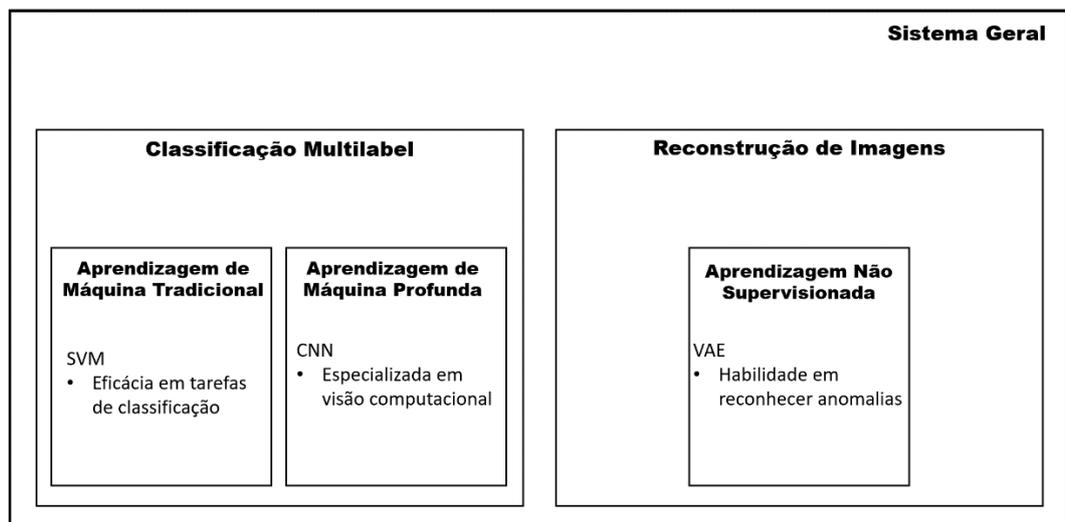
Uma destas lacunas se refere a falta de trabalhos com que faça uma comparação justa entre modelos de aprendizagem de máquina que façam a detecção

de vazamento de óleo. Para ser considerada justa, a comparação entre os modelos deveria ser realizada utilizando os mesmos banco de dados e métricas de avaliação.

Com o objetivo de conhecer melhor os padrões dos look-alikes e, conseqüentemente, ensinar a máquina a distingui-los de vazamentos de óleo de forma mais robusta, Nunes et al. (2023) construíram uma CNN que faz a multiclassificação de imagens SAR dentre as classes de look-alike. O trabalho elaborado por Conceição et al. (2021) vai além e conecta uma multiclassificação baseada num modelo de random forest a um outro modelo detector de vazamento de óleo. O princípio que motiva o treinamento da multiclassificação e a integração dos modelos é que mais informação evite falsos diagnósticos.

Dadas as lacunas identificadas na literatura, esta dissertação visa contribuir para a melhoria da detecção de vazamentos de óleo de maneira automática. O foco principal será na identificação de look-alikes, que ainda representam um grande desafio na área. A singularidade deste estudo está na utilização de modelos de que se destacam na tarefa de classificação por motivos e estilos distintos e na sugestão da combinação destes modelos na construção de um sistema robusto, como mostra a Figura 11.

Figura 11 - Sistema geral de detecção de vazamento



Fonte: Autor (2024)

Uma vez que trabalhos científicos recentes da área foram revisados, se optou por três modelos já destacados nesta seção para fazer parte do sistema. O primeiro

método construído é baseado no SVM, um modelo de aprendizagem de máquina tradicional, de grande eficácia em tarefas de classificação.

O segundo método, em contrapartida, leva em consideração a CNN que é um modelo de aprendizagem de máquina profunda e tem uma arquitetura especializada em extrair características complexas de imagens. Ambos serão empregados em um cenário de classificação multilabel, o qual apesar de introduzir certa complexidade, permite lidar com a diversidade natural de rótulos encontrados em imagens SAR.

Adiciona-se ainda o terceiro método, que está fundamentado no modelo de aprendizagem não supervisionado VAE. Capaz de processar imagens não rotuladas, o que se mostrou ser um gap na literatura, esse tipo de modelo é reconhecido pela sua habilidade em reconhecer anomalias com a vantagem de lidar com as incertezas por meio de distribuições de probabilidade.

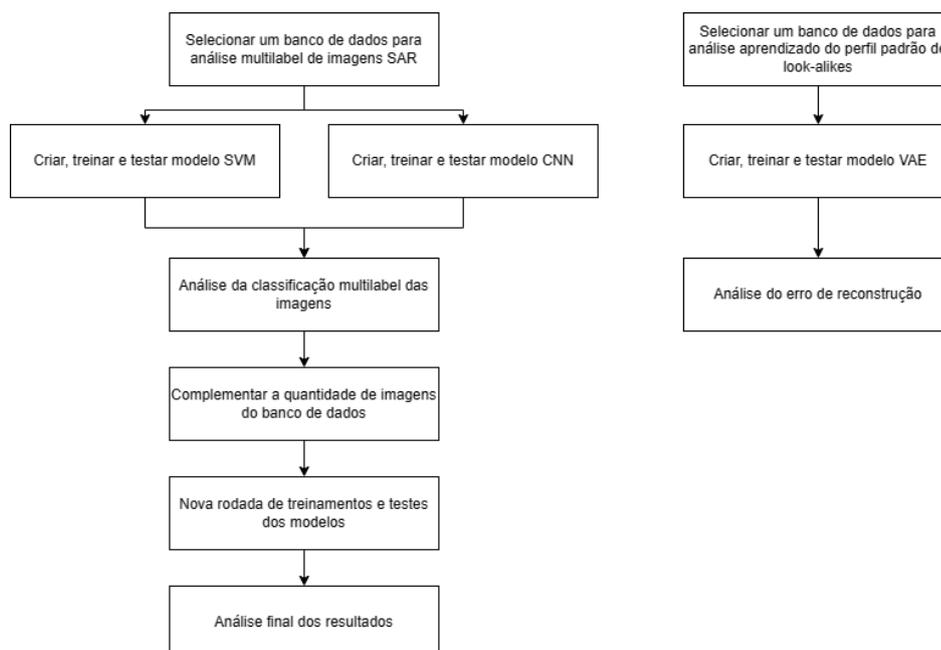
Acredita-se que os perfis distintos dos modelos se complementarão de uma maneira inovadora, e quando atuarem de maneira sinérgica alcançarão resultados mais precisos. Isto é, ao final, o sistema geral receberá o retorno de cada um dos modelos que, quando combinados, irão reforçar uma única saída, determinando com maior precisão se é um vazamento de óleo real ou não. Espera que se repita a eficácia da combinação de abordagens já comprovada anteriormente.

3 METODOLOGIA

O propósito deste trabalho de dissertação é desenvolver métodos baseados em aprendizagem de máquina com o intuito de apoiar o solucionamento de um problema real que afeta o ambiente marinho e diferentes níveis da sociedade dependentes dele. Também é esperado que este trabalho desperte o interesse de empresas e outras instituições pela possibilidade de replicação. Pode-se considerá-lo uma pesquisa de caráter prático (MIGUEL et al., 2010). Esta pesquisa também pode ser classificada como exploratória (MIGUEL et al., 2010) pois, apesar de investigar um assunto já estudado na literatura, apresenta elementos pouco, (ou quase nunca) experimentados juntos antes. A abordagem é quantitativa, já que os dados de processamento, assim como os resultados, serão analisados e representados estatisticamente.

Especificamente, busca-se melhorar a distinção entre look-alikes e vazamentos de óleo através da criação e treinamento de modelos de aprendizagem de máquina que classificam os itens presentes em imagens. Com o intuito de expor de uma maneira clara a abordagem adotada para alcançar os objetivos propostos, a metodologia desta dissertação será apresentada através de um fluxograma (Figura 12).

Figura 12 - Fluxograma da metodologia



Após rever a literatura acadêmica e criar uma base conceitual consolidada, o primeiro passo da pesquisa foi compreender os bancos de dados que foram escolhidos para estudo. O entendimento detalhado das imagens foi essencial para realização de um pré-processamento adequado para garantir a consistência das entradas fornecidas aos modelos.

O banco de imagens SAR criado por Krestenitis et al. (2019), que contém uma diversidade de classes nas imagens, tendo presença majoritariamente da classe de vazamento de óleo, foi selecionado para a abordagem de classificação multilabel. Essa escolha dificultaria o treinamento específico em look-alike pelo modelo de reconstrução de imagens. Pela necessidade de atender às características específicas do método, o banco de dados produzido por Wang et al. (2019) foi escolhido para a abordagem não supervisionada.

O segundo passo da metodologia consistiu na construção de um modelo de classificação multilabel de imagens. Entretanto, visando a um melhor desempenho, optou-se por desenvolver mais de um modelo utilizando abordagens distintas. Um desses modelos empregou um algoritmo de SVM, conhecido por sua eficácia na separação de classes complexas. O outro foi baseado numa CNN, um tipo de arquitetura profunda altamente especializada em tarefas de visão computacional. Os modelos foram desenvolvidos utilizando a linguagem de programação Python.

Paralelamente, foi-se trabalhando no procedimento de aprendizado não supervisionado porque acredita-se na eficácia de modelos generativos em representar e reconstruir dados. Para esta estratégia se elaborou um modelo VAE, o qual recorre a distribuição de probabilidades para lidar com a incerteza subjetiva dos dados.

O próximo passo foi elaborar métricas de avaliação para mensurar e analisar os modelos construídos. Através destas métricas foi possível explorar as vantagens e limitações de cada abordagem e, dessa forma, observar qual a melhor arquitetura no contexto específico.

Depois de avaliar o quão eficazmente os dois primeiros modelos estavam classificando as imagens, o passo seguinte foi identificar possíveis áreas de melhoria. Buscando aprimorar a precisão na classificação de algumas classes, mais imagens provenientes do banco de dados de Wang et al. (2019) foram incluídas no treinamento dos modelos visto que a diversidade e a quantidade de dados utilizados no treinamento podem impactar significativamente na eficácia dos modelos de aprendizado de máquina.

Por fim, o último passo foi realizar uma análise objetiva dos modelos de classificação multilabel e comparar os resultados obtidos após a incorporação das imagens complementares e retreinamento dos modelos. Esta análise permitirá tomadas de decisões mais assertivas na performance das classificações objetivo dessa dissertação.

3.1 CONJUNTOS DE DADOS

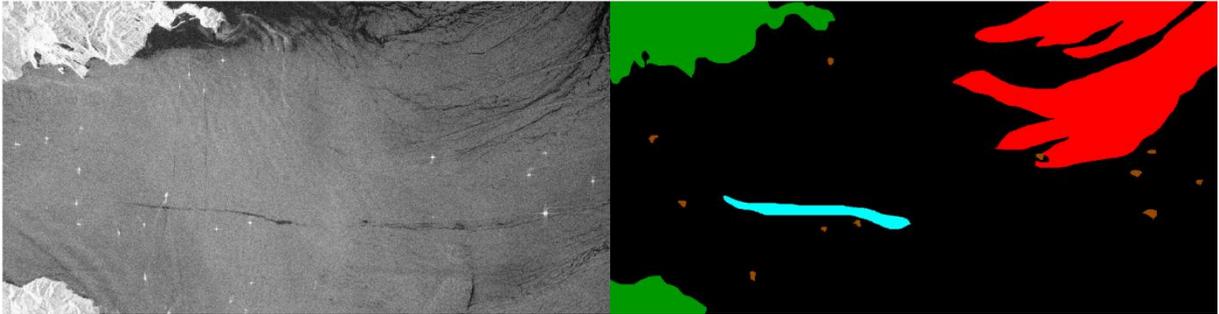
3.1.1 Oil Spill Detection Dataset

O “Oil Spill Detection Dataset” (KRESTENITIS et al., 2019) é o banco de dados aplicado na abordagem que visa a classificação multilabel desta dissertação. Trata-se de um banco de dados de acesso público, composto por imagens SAR de monitoramento do oceano (em sua maioria contendo vazamento de óleo) com áreas identificadas semanticamente por máscaras.

Imagens no formato de polarização VV (polarização vertical transmitida – polarização vertical recebida) foram selecionadas dentre tantas outras geradas pelo satélite Sentinel-1 entre 2015 e 2017 e foram disponibilizadas pela Agência Espacial Europeia (ESA – European Space Agency) através do Copernicus Open Access Hub. A confirmação do vazamento de óleo e outros dados das imagens foi feita pela Agência Europeia da Segurança Marítima (EMSA – European Maritime Safety Agency).

O “Oil Spill Detection Dataset” é formado por um conjunto de 1112 imagens, sendo 90% das imagens destinadas ao treinamento e os 10% restantes compõem o conjunto de teste do modelo. As estruturas presentes nestas imagens podem ser rotuladas em cinco classes, ilustradas na figura 13: superfície marinha, vazamento de óleo, look-alike, navio e área de costa. A superfície marinha é considerada uma classe de fundo pois está presente em todas as imagens.

Figura 13 – Imagem do Oil Spill Detection Dataset e sua máscara



Fonte: (Adaptado de KRESTENITIS et al., 2019)

As cores, ou máscaras, servem para atribuir rótulos a regiões das imagens, indicando a qual classe aquela região pertence. Na figura 13 a cor azul representa vazamento de óleo, a vermelha look-alike, a verde área de costa, a marrom navio e a preta a superfície marinha.

Seria esperado um desequilíbrio entre as classes nesse banco de imagens dada a natureza e as dimensões das estruturas de cada classe. No entanto, para torna-se robusto e representativo, o banco de dados em estudo foi construído com o objetivo de equilibrar a variedade de classes presentes nas imagens. Isso envolveu não apenas fornecer uma variedade de instâncias de cada classe, mas também considerar a diversidade de escalas das estruturas presentes nas imagens. Por meio de um esquema de extração de imagens em múltiplas escalas, as instâncias de interesse foram capturadas em diferentes tamanhos, garantindo assim uma representação abrangente das condições observadas. Desta forma, o banco de imagens reflete a complexidade e a diversidade das situações reais, mantendo uma diversidade na quantidade e na escala das imagens de cada classe.

3.2.2 TenGeoP – SARwv

Assim como o banco de dados apresentado anteriormente, o TenGeoP – SARwv (WANG et al., 2019) também é formado por imagens selecionadas dentre as disponibilizadas pela ESA através do programa Copernicus. Contudo, este outro banco de dados é formado por um conjunto de imagens apenas de look-alikes. As imagens, obtidas do Sentinel 1 em 2016, estão no formato wave mode e representam uma área com 20 x 20 km coletadas com dois ângulos de incidência alternados de

23,8° (WV1) e 36,8° (WV2) numa resolução de 5 metros no solo, obtidas na polarização padrão VV.

As imagens do TenGeoP – SARwv estão organizados originalmente em dez classes que correspondem as condições atmosféricas e oceânicas quem resultam em áreas escuras nas imagens de radar obtidas a partir dos satélites. Cada imagem pertence apenas a uma classe.

A princípio este banco de imagens só seria utilizado pela abordagem do modelo generativo. Entretanto, se fez necessário complementar o banco de dados utilizado pelos modelos na abordagem de classificação multilabel e algumas das imagens deste segundo banco de dados foram selecionadas.

O critério de seleção das imagens para complementar o banco de dados dos modelos de classificação multilabel se deu baseado com o observado em (NUNES et al., 2023), onde três dessas classes atrapalharam o desempenho de um modelo CNN treinado para a multiclassificação das imagens de look-alike, seja por estarem numa quantidade desbalanceada com relação às outras classes ou pela própria dificuldade de padronização devido ao retroespelhamento em imagens de look-alike que têm o gelo como origem. Por esse motivo, optou-se por retirar as classes “gelo marinho”, “iceberg” e “frente oceânica” do conjunto original disponível com o intuito de otimizar a seleção de imagens complementares e, conseqüentemente, os resultados desta pesquisa.

A Tabela 1 a seguir explana resumidamente as dez classes de look-alike que são consideradas pelo banco de dados, bem como o número de imagens que cada classe contém e o número total de imagens disponíveis para treinamento. As três últimas classes não estarão à disposição para seleção do banco de dados complementar levando em conta as considerações mencionadas.

Tabela 1 – Classes de look-alike que serão consideradas

#	Tipo de look-alike	Descrição	# Imagens
1	Ondas oceânicas puras	Prevalente na maioria das imagens	4900
2	Ondas de vento	Rugosidade formada pela força do vento	4797
3	Células microconectivas	Estruturas de correntes	4598
4	Células de chuva	Formas de chuva	4740
5	Manchas biológicas	Manchas formadas por seres aquáticos	4709
6	Área de vento baixo	Mar quase estacionário	2160
7	Frente atmosférica	Dois frentes de ar/vento distintas	4100

8	Gelo marinho	Água salgada congelada	4370
9	Iceberg	Água doce congelada	1980
10	Frente oceânica	Encontro de águas de massas distintas	1199
Total			33553

Fonte: Adaptado de (WANG et al., 2019)

70 imagens aleatórias deste banco de dados, sendo 10 imagens de cada uma das sete primeiras classes listadas na tabela, foram adicionadas ao “Oil Spill Detection Dataset” (KRESTENITIS et al., 2019) numa tentativa de aprimoramento dos modelos multilabel desenvolvidos para esta pesquisa. Esta decisão surge da percepção de que o número de imagens disponíveis no banco de dados pensado inicialmente para os modelos multilabel pode não ser suficiente para treiná-los efetivamente para a classe específica de look-alikes. A inclusão dessas imagens visa complementar e diversificar as amostras disponíveis para treinamento, buscando uma representação mais abrangente e robusta da variedade de características que podem ocorrer. A quantidade de imagens selecionadas foi especificada de forma a não desbalancear a proporção de imagens de cada classe no banco de dados inicial.

Já o modelo VAE utilizará exclusivamente das imagens deste banco de dados, com todas as classes, tanto para treinamento quanto para testes. Para garantir uma avaliação confiável do desempenho do modelo, 20% do total de imagens serão separadas exclusivamente para operações de validação e teste, enquanto os 80% restantes das imagens serão destinados ao treinamento do modelo.

3.1.3 Método 1 – utilizando SVM

O primeiro sistema de classificação multilabel de imagens criado para esta pesquisa usa uma abordagem baseada no SVM. O código começa importando as bibliotecas necessárias, incluindo ferramentas para manipulação de imagens, processamento de dados e métricas de avaliação. Em seguida, são definidos os caminhos para as pastas de treinamento e teste, bem como as dimensões desejadas para as imagens após o redimensionamento.

Feito isso, é projetada uma função, que no código chama-se “load_grayscale_images_from_folder”, para carregar imagens em escala de cinza da pasta de treinamento, redimensioná-las para o tamanho 256x256 pixels (altura e

largura) e normalizar seus valores de pixel. Isso cria um conjunto padronizado de imagens para análise. O carregamento das imagens de treinamento ocorre em seguida.

O carregamento das imagens tem um retorno no formato (número de imagens, altura, largura). Porém, o modelo espera um formato unidimensional, onde cada imagem estará associada a uma “feature”, sendo “feature” o número total de pixels das imagens redimensionadas resultante do produto entre a altura e a largura. Portanto, depois de carregadas, as imagens são transformadas em um formato do tipo (número de imagens, número de features) adequado para o treinamento do modelo.

Os próximos blocos do código estão relacionados as labels das imagens de treinamento. Para isso cria-se uma matriz que representa a presença ou ausência das classes específicas nas imagens. Isso é fundamental para adaptar as labels para o formato necessário, já que originalmente as labels são máscaras de segmentação semântica ou estão representadas por números que tem próprio significado.

Inicialmente uma função é projetada para associar um número de classe a uma cor correspondente da classe através de um dicionário. Depois, uma nova função é criada para obter um vetor binário que indica se cada classe está presente ou ausente em uma imagem. Por último uma função gera uma matriz desses vetores de presença para todas as imagens. Ao final, todas as labels das imagens de treinamento são carregadas.

Feito o pré-processamento do banco de imagens, os passos seguintes envolvem a elaboração do classificador SVM. Inicialmente, o Grid Search CV do pacote scikit-learn foi utilizado para encontrar os melhores hiperparâmetros para um classificador SVM multi-output. O conjunto de parâmetros que o Grid Search irá explorar foi definido com diferentes valores para o parâmetro gamma e de regularização C e diferentes kernels, como linear e RBF. Em seguida, uma instância do classificador SVM com a opção “probability=True” foi criada para calcular as probabilidades de pertencimento a cada classe e configuramos um classificador multi-output em torno dele, pois estamos lidando com múltiplas saídas. Assim, o SVM é habilitado para prever probabilidades de saída.

Finalmente, o Grid Search é executado no conjunto de treinamento, onde ele ajusta o modelo SVM com diferentes combinações de hiperparâmetros e avalia seu

desempenho usando validação cruzada, a fim de identificar a combinação que maximiza a precisão.

Com o resultado do grid search, o SVM é configurado com um kernel RBF (Radial Basis Function), o parâmetro de penalidade por erro de classificação C como sendo igual a 10 e o parâmetro gamma é definido como “scale”.

Outra parte essencial na elaboração do SVM é a criação de uma instância do classificador "MultiOutputClassifier". O "MultiOutputClassifier" da biblioteca Scikit-learn da linguagem de programação Python é uma instância que permite usar classificadores base (neste caso, um classificador SVM) para problemas de saída multi-rótulo (GÉRON, 2019). Isto acontece da seguinte maneira: O classificador SVM binário é criado e treinado para distinguir duas classes por vez, ou seja, ele considera uma classe como positiva e todas as outras classes como negativas. Gerada as previsões para cada classe individualmente, o "MultiOutputClassifier" estende essa capacidade e obtém as previsões pela combinação de todos os classificadores SVM binários, onde cada um informa se a amostra pertence ou não à classe correspondente.

Isso permite que o SVM seja usado para classificação multilabel, o que é apropriado para o problema em estudo. A partir daí, o modelo é treinado com as imagens normalizadas e as labels. Isso permite que o SVM aprenda a distinguir entre as diferentes categorias representadas nas imagens.

Após o treinamento, o sistema está pronto para fazer previsões em imagens de teste. As imagens de teste são carregadas e redimensionadas da mesma forma que as de treinamento. O classificador treinado é, então, usado para prever as categorias das imagens de teste.

Para analisar quão bem o modelo está realizando as previsões são calculadas várias métricas de avaliação de desempenho, incluindo precisão, taxa de verdadeiros positivos, F1-score, perda de Hamming e acurácia. Por fim, o código gera matrizes de confusão, uma para cada classe, em relação às classes verdadeiras das imagens de teste. As matrizes de confusão são plotadas em um formato fácil de interpretar, permitindo uma análise mais detalhada do desempenho do sistema.

3.1.4 Método 2 – utilizando CNN

Um modelo de rede neural convolucional (CNN) voltado para a classificação de imagens num cenário de múltiplas labels também foi elaborado para esta pesquisa. No início do código, são importadas as bibliotecas essenciais para construção da CNN, incluindo a TensorFlow, a NumPy, a OpenCV e a Keras (GÉRON, 2019). Adicionalmente, são estabelecidos os diretórios para as pastas de treinamento e teste do banco de dados selecionado, acompanhados das dimensões desejadas para as imagens a serem processadas, as quais serão as mesmas dimensões do código SVM.

Em seguida, é definida uma função para o processamento de imagens coloridas em RGB. Ela demonstra como carregar e preparar as imagens do conjunto de treinamento, e converte-as em um formato adequado para o modelo. Diferentemente do modelo SVM, o modelo CNN aceita o formato com as três camadas de cores RGB, não sendo necessário deixar as imagens no formato bidimensional.

Uma função para carregar as labels de treinamento também é definida similarmente a do modelo anterior. Contudo, além de criar matrizes de vetores de presença/ausência, a função do modelo CNN é mais abrangente por considerar o formato das imagens de treinamento e incluir funcionalidades específicas para criar e treinar a rede neural em questão.

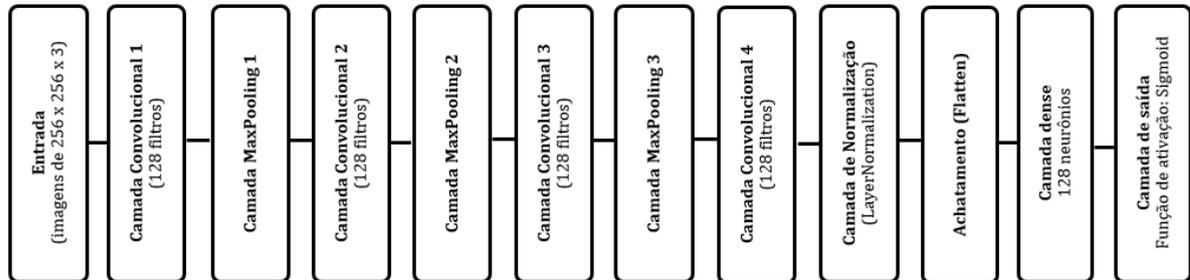
O próximo passo consiste em construir a CNN que irá receber as imagens e labels processadas previamente. Várias configurações de uma rede CNN para classificação multi-rótulo foram exploradas por meio de um grid Search. Começou-se definindo listas de valores para os parâmetros que queremos testar, como o número de filtros, o número de camadas convolucionais e de pooling, tipos de normalização e otimizadores.

Em seguida, todas as combinações possíveis desses parâmetros foram iteradas usando a função “product” do módulo “itertools”. Para cada combinação, um modelo CNN usando o Keras Sequential API com os parâmetros atuais foi construído e compilado com o otimizador e a função de perda correspondentes. Além disso, um mecanismo de parada antecipada foi configurado para evitar overfitting durante o treinamento. O conjunto de treinamento do banco de imagens foi dividido em dois outros subconjuntos (treinamento e validação) para realização do grid search.

Após a avaliação, um relatório de classificação é gerado e as matrizes de confusão são calculadas para cada classe individualmente. Os modelos foram ordenados pelo desempenho em precisão e na função perda. Este processo permitiu encontrar a configuração de modelo que melhor se adequa ao nosso conjunto de

dados. A Figura 14 explora as relações e inicia a descrição das funções de cada camada da CNN.

Figura 14 - Arquitetura da CNN



Fonte: o autor (2024)

A arquitetura da CNN é construída sequencialmente com camadas convolucionais, camadas de max pooling, camada de normalização, camada flatten e camada densa. A última camada utiliza ativação sigmoide para lidar com o problema de múltiplos rótulos. Esta ativação indica que a saída dessa camada será composta por múltiplos neurônios, onde cada neurônio está associado a uma classe específica. A ativação sigmoide é usada nesse contexto porque é adequada para problemas de classificação multilabel, onde uma amostra pode pertencer a várias classes ao mesmo tempo.

Cada neurônio na camada de saída produzirá um valor entre 0 e 1, que representa a probabilidade de a amostra pertencer à classe correspondente. Portanto, para cada amostra, será obtido um vetor de saída contendo valores de probabilidade para todas as classes, e cada classe é tratada como uma tarefa de classificação binária (pertence ou não pertence a essa classe) usando a ativação sigmoide.

Depois de definir a arquitetura, o modelo é compilado utilizando uma função de perda de entropia cruzada binária, adequada para problemas multilabel por permitir múltiplas classes para cada amostra, e um otimizador Adam. Em seguida, o modelo é treinado com base no conjunto de treinamento fornecido, em 100 épocas e com lotes de tamanho igual a trinta e dois. Durante o treinamento, vinte por cento dos dados é usada para validação.

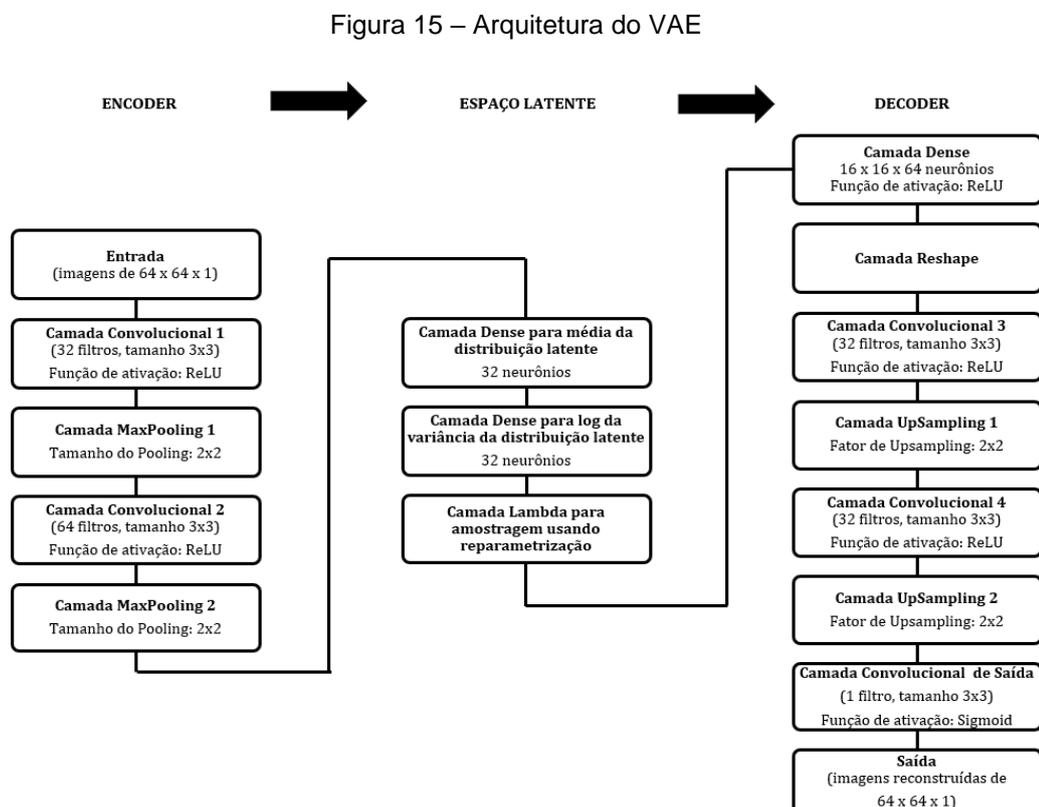
Após o treinamento do modelo, as imagens e as labels de testes são carregadas e processadas através da mesma função empregada para estes procedimentos anteriormente. Com os dados de teste prontos, o modelo realiza previsões nas imagens de teste e arredonda os resultados para 0 ou 1 para que se

alinhem com o formato dos rótulos. De modo equivalente ao modelo SVM, as métricas de avaliação são calculadas para avaliar o desempenho do modelo.

3.1.5 Método 3 – utilizando VAE

Com o intuito de explorar as vantagens de um modelo generativo não supervisionado, um modelo VAE também foi construído. O foco deste modelo é na detecção de vazamentos de óleo por meio do erro, uma vez que ele foi projetado para usar imagens de look-alikes.

Assim sendo, em situações em que ocorre um erro de reconstrução significativamente alto, podemos inferir que a imagem em questão não é um look-alike e, possivelmente, representa um vazamento de óleo real. Essa abordagem se beneficia da capacidade do VAE de discernir anomalias com base no desvio da reconstrução esperada, proporcionando uma ferramenta eficaz para a identificação de eventos adversos, como vazamentos de óleo. A figura 15 a seguir ilustra as camadas do modelo VAE construído para essa dissertação.



Primeiramente, o código importa as bibliotecas necessárias, como “os” para manipulação de arquivos, “numpy” para manipulação de arrays, “PIL” para trabalhar com imagens, e módulos das bibliotecas “Scikit-Learn” para dividir um conjunto de dados em subconjuntos de treinamento e teste e “Keras” para construir e treinar o modelo (GÉRON, 2019).

Em seguida, todas as imagens do banco de dados “TenGeoP – SARwv” (WANG et al., 2019) são carregadas, convertidas para a escala de cinza e redimensionadas para o tamanho de 64x64 pixels. Além de reduzir dimensionalidades, lidar com apenas um canal de cor permite que o modelo foque na textura das imagens, uma vez que cor não é um fator tão importante para o caso em estudo. Depois de carregadas, as imagens são pré-processadas. Isso inclui a conversão das listas de imagens para arrays numpy, a divisão do conjunto de dados em conjuntos de treinamento e teste, a normalização dos valores de pixel para o intervalo [0, 1], e a adição de uma dimensão de canal para as imagens em escala de cinza.

A partir daí é que a arquitetura do VAE começa a ser construída. Inicialmente, a função “sampling”, que desempenha um papel crucial na amostragem da distribuição latente, é definida. Essa função utiliza a técnica de reparametrização de Monte Carlo, que é essencial para treinar o modelo de maneira eficaz. Ela recebe como argumentos a média (z_mean) e o logaritmo da variância (z_log_var) da distribuição latente. Ao incorporar uma amostra aleatória da distribuição normal padrão, a função produz uma representação estocástica da variável latente. Esse processo de reparametrização é fundamental para tornar o treinamento do VAE viável através do gradiente descendente estocástico.

Logo depois, é criada a camada de entrada “input_img” que irá alimentar o VAE. A forma da imagem para entrada é especificada como ($image_size[0]$, $image_size[1]$, 1), indicando as dimensões da imagem (altura, largura, canal). Neste caso, o canal é definido como 1, já as imagens estarão em escala de cinza. Essa camada de entrada é crucial para estabelecer a estrutura do modelo, indicando como os dados serão recebidos e processados durante o treinamento.

Posteriormente, é definida a arquitetura do encoder do VAE por meio de uma sequência de camadas convolutivas e de pooling. O objetivo é aprender características importantes das imagens de entrada. A escolha de uma arquitetura baseada em CNNs se justifica na capacidade dessas redes de extrair características

relevantes, apresentar robustez para variação de escala e deformações nas imagens, reduzir a dimensionalidade das imagens de forma intrínseca, ser computacionalmente eficientes, flexíveis e ter um histórico de sucesso em tarefas de reconstrução de imagens (AL BATAINEH et al., 2023; BHATT et al., 2021; DONG et al., 2014).

A primeira camada “Conv2D” aplica 32 filtros de convolução de tamanho 3x3, seguida de uma função de ativação ReLU. A camada “MaxPooling2D” reduz a dimensionalidade da imagem pela metade. A segunda camada “Conv2D” tem 64 filtros e é seguida por outra camada “MaxPooling2D”.

Após as camadas convolutivas, a saída é achatada para um vetor unidimensional usando “Flatten()”. E duas camadas densas (Dense) são adicionadas para estimar a média (z_mean) e o logaritmo da variância (z_log_var) da distribuição latente. Estes valores serão utilizados para a amostragem na distribuição latente.

Uma camada “Lambda” é usada para realizar a amostragem da distribuição latente usando a função definida anteriormente (“sampling”). Essa camada recebe como entrada as estimativas da média (z_mean) e log da variância (z_log_var), e produz a amostra da distribuição latente (z).

Finalmente, todas essas camadas são agrupadas para formar o encoder completo. O modelo recebe uma imagem de entrada ($input_img$) e retorna três saídas: a média (z_mean), o log da variância (z_log_var), e a amostra da distribuição latente (z).

De maneira similar a parte do decoder do VAE é construída. A princípio, é criada uma camada de entrada que receberá a amostra da distribuição latente. Esta entrada é essencial para gerar as reconstruções das imagens. A forma da entrada é especificada como (32,), indicando a quantidade de dimensões que a amostra latente possui.

Em seguida, utilizamos camadas densas e operações de reshaping para expandir a amostra latente para um formato adequado para a reconstrução da imagem. A camada densa é responsável por aumentar a dimensionalidade, e a camada de reshaping transforma o vetor resultante em um tensor 4D com dimensões (16, 16, 64, 1) considerando ($batch_size$, altura, largura, canais) da imagem.

Posteriormente, aplicamos camadas convolutivas e operações de upsampling para aprender representações mais complexas e espaciais do espaço latente. As camadas convolutivas ajudam a capturar características significativas, enquanto as operações de upsampling aumentam as dimensões espaciais do tensor.

A camada de saída (“decoded”) é a última etapa do decoder. Utiliza uma camada convolutiva com ativação sigmoid para gerar a reconstrução final da imagem. A ativação sigmoid garante que os valores de saída estejam no intervalo de 0 a 1, tornando a saída compatível com imagens em escala de cinza. Por último, o modelo do decoder é instanciado, recebendo as amostras latentes como entrada e produzindo as reconstruções das imagens como saída.

O modelo completo do VAE é construído conectando o encoder e o decoder. Uma função de perda personalizada é definida como a média da soma da entropia cruzada binária, escolhida neste caso com base na natureza dos dados de entrada, na capacidade de avaliar a similaridade entre as distribuições de probabilidade e na eficácia geral para treinar o modelo na tarefa de reconstrução das imagens, e a divergência de Kullback-Leibler. Isso é feito para treinar o VAE de modo que ele seja capaz de gerar imagens que se assemelhem às originais, ao mesmo tempo em que aprende uma representação compacta na distribuição latente.

A entropia cruzada binária (BCE) é uma medida específica para avaliar a similaridade entre as distribuições de probabilidade das imagens originais e das reconstruções. No contexto do VAE, a BCE é frequentemente utilizada como parte da função de perda durante o treinamento. Ela compara a distribuição de probabilidade dos pixels individuais nas imagens reconstruídas com a distribuição de probabilidade dos pixels nas imagens originais. Uma baixa pontuação nessa métrica indica uma boa correspondência entre as distribuições, destacando a capacidade do modelo em replicar a complexidade estatística dos dados.

A divergência de Kullback-Leibler (KL) é uma medida crucial para a regularização do espaço latente do VAE. Ela quantifica a diferença entre a distribuição latente aprendida pelo modelo e uma distribuição latente de referência, geralmente uma distribuição normal padrão. Durante o treinamento, a minimização da KL assegura que o espaço latente seja bem organizado e coerente, promovendo uma representação compacta e estruturada das variações nos dados de entrada. A divergência KL é um componente crucial da função de perda do VAE e ajuda a garantir que o modelo não apenas reproduza dados de entrada, mas também aprenda uma representação significativa e generalizável.

Finalmente, o modelo VAE é compilado com o otimizador Adam e é treinado usando o conjunto de treinamento. O método “fit” é utilizado para ajustar os pesos do modelo com base nos dados de treinamento. O treinamento ocorrerá ao longo de 150

épocas e o modelo será atualizado a cada lote de 32 amostras. Além disso, uma fração de 20% dos dados de treinamento serão reservados para validação durante o treinamento. Isso é útil para monitorar o desempenho do modelo em dados não vistos durante o treinamento e evitar o overfitting.

Após ser treinado, o modelo é programado para gerar reconstruções das imagens de teste. Em seguida, é realizada a avaliação do desempenho do modelo por meio do cálculo do erro médio de reconstrução. Essa métrica mede a diferença média entre as imagens originais e as imagens reconstruídas pelo VAE, fornecendo uma informação de quão bem o modelo é capaz de reproduzir as entradas originais.

4 ANÁLISE E DISCUSSÃO DOS RESULTADOS

Dadas as métricas de avaliação definidas na seção 2.2.5, pode-se avaliar e interpretar os resultados obtidos por cada um dos modelos de aprendizagem de máquina em estudo. A Tabela 2 apresenta os valores obtidos pelo modelo SVM e pelo modelo CNN e permite uma comparação entre eles.

Tabela 2 – Métricas para avaliação dos modelos

	SVM	CNN
Average Precision	68%	88%
Average Recall	61%	62%
Average F1-score	66%	65%
Hamming Loss	18%	15%
Subset Accuracy	26%	38%

Fonte: O Autor (2024)

O modelo CNN evidencia sua vantagem significativa em fazer previsões mais precisas se comparado ao modelo SVM. Também chama a atenção a porcentagem correspondente a acurácia de subconjunto do modelo CNN, sinalizando uma capacidade superior em prever todas as labels corretamente. Outra comparação positiva do modelo é a taxa menor de rótulos incorretos correspondente a métrica de perda de Hamming.

Comparando o desempenho dos modelos testados em termos especificamente de falsos positivos, também observamos uma certa vantagem do modelo que usa CNN em relação ao SVM. Dentre as métricas testadas, a mais sensível aos falsos positivos é a “Average Precision”. Quanto mais falsos positivos forem identificados, menor será o valor resultante da métrica.

O modelo SVM alcançou uma “Average Precision” de 68% que, embora ainda seja uma métrica respeitável, sugere que ela pode gerar um número um pouco maior de falsos positivos em comparação com os 88% obtidos pelo modelo CNN, métrica que destaca a eficácia na gestão desses erros pelo segundo método testado.

Logo, pode-se concluir que, mesmo com resultados similares da taxa média de valores verdadeiramente positivos e do F1-score, o modelo CNN indica uma melhor capacidade de identificar as labels do contexto específico da pesquisa visto que apresenta resultados superiores ao modelo SVM na maioria das métricas. Uma possível explicação para o desempenho superior do modelo CNN em comparação

com o modelo SVM na maioria das métricas é a capacidade intrínseca das redes neurais convolucionais profundas em lidar com características complexas e hierárquicas nos dados.

Entretanto, a eficácia do modelo se destaca ao examinar a precisão observada em cada classe individual, conforme revelado pelas matrizes de confusão. Esses dados são detalhados na Tabela 3.

Tabela 3 - Proporção de acertos do modelo SVM

Classe	Proporção de acertos
Vazamento de óleo	82%
Look-alike	65%
Navio	71%
Área de costa	89%

Fonte: O Autor (2024)

Em conjuntos de dados que abrangem diversas classes, é comum que a precisão média seja influenciada por classes menos representadas ou mais desafiadoras de discriminar. No entanto, se o modelo está consistentemente obtendo altas taxas de acerto para classes específicas, isso sugere sua capacidade de efetivamente discernir entre diferentes padrões e atributos associados a cada classe. Essa habilidade é crucial para a sua eficácia em situações práticas.

Contudo, esperavam-se valores percentuais maiores do que os apresentados inicialmente pelas métricas de avaliação. E quando os resultados das matrizes de confusão de cada classe, que estão anexadas nos apêndices, são observados, percebe-se que a classe que retorna os menores número de acertos de classificação é justamente a classe de look-alike.

Assumindo que isso seja a causa dos percentuais abaixo do esperado, novos testes foram feitos. A proposta para os testes foi aumentar o número de dados de treinamento da classe look-alike com a intenção de melhorar o reconhecimento do padrão dessa classe pelo modelo utilizando o banco de dados complementar para isso. Teve-se o cuidado de não incluir uma quantidade desproporcional de imagens para não desbalancear o banco de dados.

Adicionar todas as imagens do banco de dados complementar ao banco principal faria com que tivéssemos uma quantidade muito grande de imagens da classe look-alike para treinamento e isso poderia trazer efeitos que queremos evitar,

como overfitting. Para lidar com isso, foi acrescentada ao modelo uma arquitetura onde apenas dez imagens aleatórias de cada uma das sete classes do banco de dados complementar são selecionadas.

No total, 70 imagens de look-alike são incluídas ao banco de dados principal para retreinamento dos modelos, conforme descrito na seção 3.2.2. Originalmente, o conjunto de treinamento do banco de dados principal consiste em 1002 imagens, das quais 569 contém a classe look-alike, representando cerca de 57% do total das imagens de treinamento. Quando acrescentamos as 70 imagens do banco de dados complementar, essa porcentagem passa para 59,6%, mantendo um balanceamento na distribuição das classes semelhante ao apresentado anteriormente.

As imagens que foram acrescentadas ao banco de dados principal contêm apenas a classe look-alike. Os vetores binários associados a cada uma destas imagens foram feitos automaticamente através de código no Python. A coluna associada a classe de look-alike era a única que recebia valores iguais a 1, todas as outras quatro colunas receberam valores iguais a 0.

Após o retreinamento, os modelos foram avaliados pelas métricas novamente. A proporção de acertos retirados das matrizes de confusão depois desses ajustes, assim como os resultados obtidos previamente são apresentados nas Tabelas 4 e 5 para comparação.

Tabela 4 – Proporção de acertos do modelo SVM

	Treinamento inicial	Retreinamento
Vazamento de óleo	82%	82%
Look-alike	65%	54%
Navio	71%	67%
Área de costa	89%	75%

Fonte: O Autor (2023)

Tabela 5 – Proporção de acertos do modelo CNN

	Treinamento inicial	Retreinamento
Vazamento de óleo	81%	75%
Look-alike	65%	68%
Navio	68%	60%
Área de costa	90%	70%

Fonte: O Autor (2023)

Ao avaliar os valores das tabelas, nota-se que há um ganho de alguns poucos pontos percentuais com o retreinamento do modelo para a label de look-alike, incluindo imagens do banco de dados complementar. Entretanto, esta pequena melhora em pontos percentuais custa a piora no desempenho das outras labels.

Um ponto interessante que emerge na busca por interpretar os resultados das matrizes de confusão é a constatação de que de uma forma geral o modelo SVM retorna melhores resultados para cada label individualmente. Ao comparar o desempenho dos dois modelos, percebe-se que o modelo SVM tende a produzir proporções de acertos mais altas para cada label específica em relação ao modelo CNN. Isso sugere que o modelo SVM é capaz de realizar previsões parciais mais assertivas e confiáveis.

A qualidade das reconstruções e o desempenho do modelo VAE também foram analisados. Para efeitos de comparação, as funções perda que contribuíram na regularização do espaço latente no treinamento do modelo foram adicionalmente calculadas para o modelo como um todo usando as imagens do conjunto de teste. Os valores médios resultantes são apresentados na tabela 5.

Tabela 6 - Métricas de avaliação do modelo VAE

Erro médio de reconstrução	0.0001015
Entropia Cruzada Binária	1.7015
Divergência de Kullback-Leibler (kl_loss)	0.4263

Fonte: O Autor (2024)

Um erro médio de reconstrução de aproximadamente 1.015×10^{-4} indica que, em média, as reconstruções geradas pelo modelo VAE são muito precisas em relação às imagens originais do conjunto de teste. Este valor é bastante baixo e próximo de zero, sugerindo uma reconstrução de alta qualidade.

As outras duas perdas médias validam o desempenho satisfatório do modelo. Os valores relativamente baixos da entropia cruzada binária e da divergência de Kullback-Leibler sugerem que o modelo VAE está conseguindo regularizar bem o espaço latente e é capaz de gerar reconstruções com boa qualidade.

A fim de ilustrar a aplicação da abordagem proposta, foram conduzidos testes adicionais utilizando uma imagem contendo óleo e outra com look-alike do banco de dados (KRESTENITIS et al., 2019). O resultado para efeito de comparação está apresentado na tabela 6 a seguir.

Tabela 7 – Teste adicional VAE

	Imagem com óleo	Imagem com look-alike
Erro médio de reconstrução	0.65	0.41

Fonte: O Autor (2024)

Ao comparar o desempenho do modelo VAE na reconstrução dessas imagens, observa-se uma diferença nos valores resultantes. Especificamente, o erro médio de reconstrução para a imagem com óleo é consideravelmente maior do que para a imagem com look-alike. Já que o modelo foi treinado apenas com imagens sem óleo é razoável, portanto, esperar que ele tenha dificuldade em reconstruir imagens com óleo.

O resultado de erro médio de reconstrução obtido para as imagens com look-alike também se apresentou, de certa forma, como esperado. Este valor pode indicar um desequilíbrio entre o conjunto de dados utilizado para treinamento do modelo e o do caso deste teste.

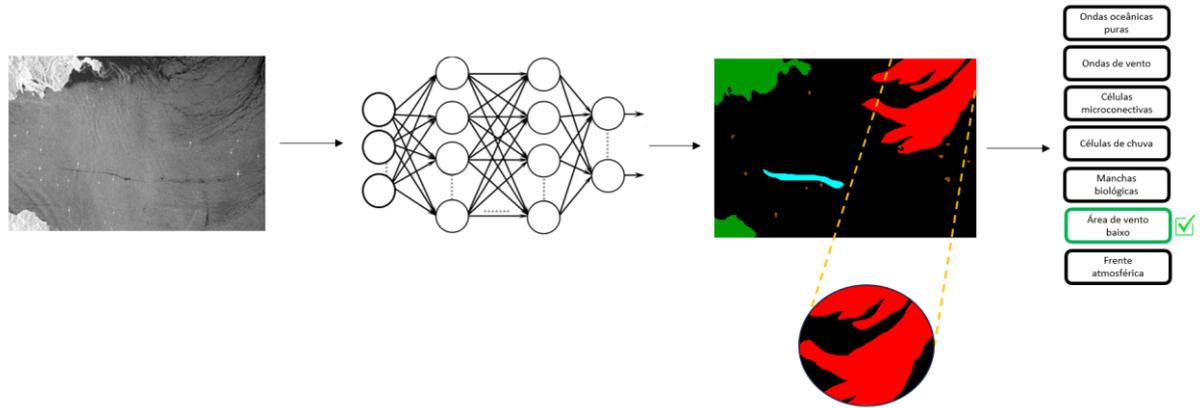
Quando o modelo é exposto a imagens fora do conjunto de treinamento, ele pode não ser capaz de reconstruí-las com a mesma precisão que as imagens nas quais foi treinado. O valor observado pode ser atribuído à necessidade de lidar com as particularidades específicas dos bancos de imagens. Por exemplo, o fato de as imagens deste novo teste estarem em um formato de qualidade diferente, como TIFF, pode ter sido a causa deste resultado.

Logo, é importante levar em consideração que os modelos de aprendizado de máquina, incluindo os VAEs, têm limitações em sua capacidade de generalização para dados fora do conjunto de treinamento. E, ao avaliar o desempenho do modelo em novos tipos de imagens, é esperado encontrar um aumento no erro de reconstrução.

No entanto, os resultados apresentados pelos três modelos ainda indicam que eles estão aptos para integrar um sistema abrangente de detecção de óleo, como o representado na figura 16. Cada modelo possui suas particularidades, e quando

combinados produzem uma saída mais estável que reduz a probabilidade de ocorrência de falsos positivos.

Figura 16 - Sistema completo



Fonte: O Autor (2024)

O conhecimento aprimorado sobre look-alikes e a capacidade de identificá-los automaticamente por meio de modelos de aprendizagem de máquina representam uma contribuição substancial na identificação de vazamentos reais de óleo. A informação de um sistema integrado e mais aprofundado acerca de look-alike como o ilustrado na figura 16 pode proporcionar além de uma detecção precisa, economia de recursos, respostas rápidas e proteção mais eficiente de ambientes marinhos.

5 CONCLUSÃO

A aprendizagem de máquina é um campo de conhecimento em pleno desenvolvimento e bastante significativo numa era de transição digital como a que estamos passando. A aplicação dessas tecnologias que são capazes de lidar com volumes massivos de dados em busca de resultados otimizados abre novas oportunidades em uma ampla gama de domínios. A indústria de óleo e gás, por exemplo, deseja garantir a confiabilidade e diminuir riscos relacionados ao derramamento de óleo nos oceanos e tem no monitoramento e controle destes ambientes complexos um fator desafiador. Assim sendo, o objetivo maior desta pesquisa é aprimorar abordagens distintas, que se complementem mutualmente num sistema integrado, visando reduzir a incidência de falsos positivos na detecção automática de vazamentos de óleo.

A primeira abordagem busca aprimorar a capacidade de identificação e diferenciação entre incidentes de derramamento de óleo e look-alikes por meio da classificação multi-rótulo de imagens de monitoramento obtidas por radar de satélites. Em problemas de classificação multi-rótulo, cada amostra de dados pode pertencer a várias classes ao mesmo tempo, em oposição a problemas de classificação multiclasse, onde cada amostra pertence a uma única classe exclusiva.

Para isso, exploramos dois modelos de aprendizagem de máquina: o SVM, um método mais convencional, e a CNN, rede neural que é amplamente utilizada em tarefas que envolvem imagens. Ao avaliar os resultados obtidos pelos métodos elaborados com base nesses modelos, algumas considerações surgiram e permitiram uma análise crítica do desempenho deles.

Inicialmente, os resultados revelaram que a CNN demonstrou uma vantagem significativa em relação ao SVM, apresentando previsões mais precisas e uma capacidade superior de identificar múltiplos rótulos. Isso ficou evidenciado pelas métricas de avaliação, incluindo a Acurácia de Subconjunto e a taxa de Hamming Loss.

No entanto, a análise mais aprofundada revelou que, apesar da superioridade geral da CNN, o SVM tende a produzir proporções de acertos mais altas para cada label específica. Essa descoberta levanta a questão da capacidade do SVM de

realizar predições parciais mais assertivas e confiáveis para cada classe individualmente.

Adicionalmente, foram conduzidos experimentos de retreinamento dos modelos, incluindo imagens do banco de dados complementar para a classe "Look-alike." Embora tenha havido um leve aumento na proporção de acertos para essa classe, observou-se uma piora no desempenho dos outros rótulos, sem impactos significativos nas métricas de avaliação globais dos modelos.

Por outro lado, a segunda abordagem direcionou-se à implementação de técnicas de aprendizado não supervisionado, mais especificamente ao modelo generativo VAE. Ao aprender representações latentes probabilísticas dos dados, o VAE revela-se eficaz na detecção de anomalias.

O treinamento do VAE, com foco na minimização do erro de reconstrução, resultou em um modelo capaz de reproduzir fielmente as imagens originais de look-alike, como evidenciado pelos valores obtidos nas suas métricas avaliativas. Sob o pressuposto de que essas métricas não apenas quantificam a precisão das reconstruções, mas também demonstram a eficácia do modelo na regularização do espaço latente, pôde-se garantir uma representação significativa dos dados.

Considerando que o modelo VAE construído para esta dissertação foi treinado com imagens de look-alikes, quando encontrar uma situação em que o erro de reconstrução é significativamente alto, isso pode indicar que a imagem apresenta padrões não esperados, sugerindo a presença de um possível vazamento de óleo. Entretanto, essa não é uma resposta precisa de detecção de óleo pois trata-se apenas um indício.

Para se tornar uma informação mais segura, a saída desse modelo deveria ser avaliada simultaneamente com o que é retornado de outros modelos de aprendizagem de máquina. Este é mais um aspecto abordado nesta dissertação, que propõe a construção de um produto tecnológico, como um sistema ou aplicativo. A finalidade é facilitar a identificação rápida e precisa de potenciais vazamentos de óleo, reduzindo as chances de falsos positivos devido a semelhanças visuais, por meio da complementaridade de abordagens e modelos integrados no sistema.

5.1 LIMITAÇÕES

Apesar dos esforços dedicados à melhoria dos índices de acurácia na classificação multi-rótulo, este estudo enfrentou algumas limitações que merecem serem discutidas. Ao tentar aprimorar o treinamento do modelo, optou-se por expandir o banco de dados, incorporando imagens adicionais da classe look-alike. No entanto, a limitação crítica foi que o conjunto de dados complementar consistia exclusivamente de look-alikes. A ausência de outras categorias representa uma falta de diversidade nos dados de treinamento, o que poderia afetar o desempenho do modelo na classificação das outras categorias.

Outra limitação significativa foi a falta de informações detalhadas sobre a origem dos look-alikes nas imagens do conjunto de dados principal. O fenômeno look-alike pode surgir de várias condições oceânicas e atmosféricas, como nuvens, áreas de frente de vento, entre outros. Essa falta de detalhes também pode ter dificultado o processo de treinamento do modelo quando adicionamos imagens do outro banco uma vez que uma classe de look-alike poderia estar sendo priorizada.

Essas limitações enfatizam a complexidade das tarefas de classificação multi-rótulo, especialmente quando se lida com classes específicas, como “look-alikes”, que podem ser originadas de múltiplos fenômenos. A falta de informações detalhadas sobre a origem desses look-alikes e a falta de diversidade nos dados de treinamento representaram desafios significativos na busca por melhorar a assertividade do modelo.

É importante reconhecer que essas limitações oferecem oportunidades para pesquisas futuras, incluindo a exploração de estratégias mais equilibradas para expansão do banco de dados e a busca por fontes de dados mais abrangentes e detalhadas que possam fornecer informações contextualizadas sobre as categorias em questão. A superação dessas limitações pode contribuir para o desenvolvimento de modelos mais robustos e precisos no futuro.

5.2 SUGESTÕES PARA TRABALHOS FUTUROS

Esta pesquisa também identificou várias oportunidades de pesquisa que podem aprofundar o entendimento e aprimorar a eficácia da detecção de vazamentos de óleo e look-alikes em imagens de monitoramento de satélite. Considerando as limitações

e desafios encontrados, as seguintes sugestões são oferecidas como direções promissoras para trabalhos futuros.

Para melhorar ainda mais a eficácia dos modelos na detecção de vazamentos de óleo e de look-alikes em imagens de monitoramento por satélite, uma sugestão é realizar um grid search mais abrangente, explorando uma gama mais ampla de parâmetros para otimizar as arquiteturas construídas.

Uma abordagem para melhorar a precisão da classificação é a implementação de técnicas de segmentação semântica mais avançadas. O uso de diferentes técnicas de segmentação semântica permitirá uma análise mais abrangente e precisa das imagens (CRUZ-ROJAS et al., 2023), o que pode levar a informações mais detalhadas sobre a extensão e a natureza de um derramamento de óleo ou look-alikes.

A incorporação de técnicas de anisotropia é outra área promissora de pesquisa. A anisotropia pode ser usada para identificar e explorar a direcionalidade e padrões nas imagens de monitoramento. Isso pode ser particularmente útil para detectar características específicas associadas a derramamentos de óleo ou outros fenômenos (MIRANDA et al., 2023). Uma pesquisa futura pode investigar como a anisotropia pode ser integrada às abordagens existentes para melhorar a classificação.

Outra sugestão é a exploração de abordagens baseadas em modelos informados pela física (physics-informed models). Esses modelos podem incorporar conhecimentos físicos sobre o comportamento dos vazamentos com técnicas avançadas de aprendizado de máquina, sendo possível desenvolver modelos mais robustos e eficazes para detecção e monitoramento de vazamentos em imagens SAR.

REFERÊNCIAS

- ABOU SAMRA, R. M.; ALI, R. R. Monitoring of oil spill in the offshore zone of the Nile Delta using Sentinel data. **Marine Pollution Bulletin**, v. 179, p. 113718, jun. 2022.
- ALZUBAIDI, L. et al. Review of deep learning: concepts, CNN architectures, challenges, applications, future directions. **Journal of Big Data**, v. 8, n. 1, p. 53, 31 mar. 2021.
- AMRI, E. et al. Offshore Oil Slick Detection: From Photo-Interpreter to Explainable Multi-Modal Deep Learning Models Using SAR Images and Contextual Data. **Remote Sensing**, v. 14, n. 15, 1 ago. 2022.
- ANSELL, D. V. et al. A Review of the Problems Posed By Spills of Heavy Fuel Oils. **International Oil Spill Conference Proceedings**, v. 2001, n. 1, p. 591–596, 1 mar. 2001.
- BENGIO, Y.; COURVILLE, A.; VINCENT, P. Representation learning: A review and new perspectives. **IEEE Transactions on Pattern Analysis and Machine Intelligence**, v. 35, n. 8, p. 1798–1828, 2013.
- BHANGALE, U. et al. High performance GPU computing based approaches for oil spill detection from multi-temporal remote sensing data. **Remote Sensing of Environment**, v. 202, p. 28–44, 1 dez. 2017.
- BOWMAN, S. R. et al. **Generating Sentences from a Continuous Space**. [s.l.: s.n.]. BROCKI, L.; CHUNG, N. C. **Concept saliency maps to visualize relevant features in deep generative models**. Proceedings - 18th IEEE International Conference on Machine Learning and Applications, ICMLA 2019. **Anais...**Institute of Electrical and Electronics Engineers Inc., 1 dez. 2019.
- CERVANTES, J. et al. A comprehensive survey on support vector machine classification: Applications, challenges and trends. **Neurocomputing**, v. 408, p. 189–215, 30 set. 2020.
- CHENG, G. et al. Remote Sensing Image Scene Classification Meets Deep Learning: Challenges, Methods, Benchmarks, and Opportunities. **IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing**, v. 13, p. 3735–3756, 2020.
- CHENG, G.; ZHOU, P.; HAN, J. Learning Rotation-Invariant Convolutional Neural Networks for Object Detection in VHR Optical Remote Sensing Images. **IEEE Transactions on Geoscience and Remote Sensing**, v. 54, n. 12, p. 7405–7415, dez. 2016.
- CHICCO, D.; JURMAN, G. The advantages of the Matthews correlation coefficient (MCC) over F1 score and accuracy in binary classification evaluation. **BMC Genomics**, v. 21, n. 1, 2 jan. 2020.

CONCEIÇÃO, M. R. A. et al. SAR Oil Spill Detection System through Random Forest Classifiers. **Remote Sensing**, v. 13, n. 11, p. 2044, 22 maio 2021.

DE LA HUZ, R.; LASTRA, M.; LÓPEZ, J. Other Environmental Health Issues: Oil Spill. Em: **Encyclopedia of Environmental Health**. [s.l.] Elsevier, 2019. p. 792–796.

DHAVALIKAR, A. S.; CHOUDHARI, P. C. **CLASSIFICATION OF OIL SPILLS AND LOOK-ALIKES FROM SAR IMAGES USING BAG OF VISUAL WORDS METHOD OF FEATURE EXTRACTION**. International Geoscience and Remote Sensing Symposium (IGARSS). **Anais...**Institute of Electrical and Electronics Engineers Inc., 2021.

GÉRON, A. **Hands-On Machine Learning with Scikit-Learn, Keras, and TensorFlow SECOND EDITION Concepts, Tools, and Techniques to Build Intelligent Systems**. [s.l.: s.n.]. Disponível em: <<http://oreilly.com>>.

GONG, D. et al. **Memorizing normality to detect anomaly: Memory-augmented deep autoencoder for unsupervised anomaly detection**. Proceedings of the IEEE International Conference on Computer Vision. **Anais...**Institute of Electrical and Electronics Engineers Inc., 1 out. 2019a.

GONG, D. et al. **Memorizing normality to detect anomaly: Memory-augmented deep autoencoder for unsupervised anomaly detection**. Proceedings of the IEEE International Conference on Computer Vision. **Anais...**Institute of Electrical and Electronics Engineers Inc., 1 out. 2019b.

GOODFELLOW, I.; BENGIO, Y.; COURVILLE, A. **Deep Learning** . Disponível em: <<http://www.deeplearningbook.org>>. Acesso em: 12 maio. 2023.

GUO, Y.; ZHANG, H. Z. Oil spill detection using synthetic aperture radar images and feature selection in shape space. **International Journal of Applied Earth Observation and Geoinformation**, v. 30, n. 1, p. 146–157, 2014.

HU, F. et al. Transferring Deep Convolutional Neural Networks for the Scene Classification of High-Resolution Remote Sensing Imagery. **Remote Sensing**, v. 7, n. 11, p. 14680–14707, 5 nov. 2015.

HU, J. et al. Squeeze-and-Excitation Networks. **IEEE Transactions on Pattern Analysis and Machine Intelligence**, v. 42, n. 8, p. 2011–2023, 1 ago. 2020.

HUBY, A. A.; ALUBADY, R.; SAGBAN, R. **Oil Spill Segmentation from SAR Images Using Deep Neural Networks**. 2022 International Symposium on Multidisciplinary Studies and Innovative Technologies (ISMSIT). **Anais...**IEEE, 20 out. 2022. Disponível em: <<https://ieeexplore.ieee.org/document/9932853/>> Acesso em: 12 maio. 2023.

JIN, H. et al. Integrated leakage detection and localization model for gas pipelines based on the acoustic wave method. **Journal of Loss Prevention in the Process Industries**, v. 27, p. 74–88, jan. 2014.

JORDAN, M. I.; MITCHELL, T. M. Machine learning: Trends, perspectives, and prospects. **Science**, v. 349, n. 6245, p. 255–260, 17 jul. 2015.

KECMAN, V. Support Vector Machines – An Introduction. Em: [s.l: s.n.]. p. 1–47.

KHALID, H.; WOO, S. S. **OC-FakeDect: Classifying Deepfakes Using One-class Variational Autoencoder**. 2020 IEEE/CVF Conference on Computer Vision and Pattern Recognition Workshops (CVPRW). **Anais...IEEE**, jun. 2020.

KRESTENITIS, M. et al. Oil Spill Identification from Satellite Images Using Deep Neural Networks. **Remote Sensing**, v. 11, n. 15, p. 1762, 26 jul. 2019.

LECUN, Y. et al. Backpropagation Applied to Handwritten Zip Code Recognition. **Neural Computation**, v. 1, n. 4, p. 541–551, dez. 1989.

LECUN, Y.; BENGIO, Y.; HINTON, G. Deep learning. **Nature**, v. 521, n. 7553, p. 436–444, 28 maio 2015.

LECUN, Y.; KAVUKCUOGLU, K.; FARABET, C. **Convolutional networks and applications in vision**. Proceedings of 2010 IEEE International Symposium on Circuits and Systems. **Anais...IEEE**, maio 2010.

LI, X. et al. Line-CNN: End-to-End Traffic Line Detection With Line Proposal Unit. **IEEE Transactions on Intelligent Transportation Systems**, v. 21, n. 1, p. 248–258, jan. 2020.

LITVINENKO, V. et al. **Global guidelines and requirements for professional competencies of natural resource extraction engineers: Implications for ESG principles and sustainable development goals**. **Journal of Cleaner Production** Elsevier Ltd, , 1 mar. 2022.

LIU, W. et al. A survey of deep neural network architectures and their applications. **Neurocomputing**, v. 234, p. 11–26, 19 abr. 2017.

LIU, X. et al. Multi-source knowledge graph reasoning for ocean oil spill detection from satellite SAR images. **International Journal of Applied Earth Observation and Geoinformation**, v. 116, 1 fev. 2023.

LIZ, H. et al. Deep learning for understanding multilabel imbalanced Chest X-ray datasets. **Future Generation Computer Systems**, v. 144, p. 291–306, 1 jul. 2023.

LOTFOLLAHI, M. et al. Deep packet: a novel approach for encrypted traffic classification using deep learning. **Soft Computing**, v. 24, n. 3, p. 1999–2012, 1 fev. 2020.

MA, L. et al. **Deep learning in remote sensing applications: A meta-analysis and review**. **ISPRS Journal of Photogrammetry and Remote Sensing** Elsevier B.V., , 1 jun. 2019.

MAIOR, C. B. S. et al. Convolutional neural network model based on radiological images to support COVID-19 diagnosis: Evaluating database biases. **PLOS ONE**, v. 16, n. 3, p. e0247839, 1 mar. 2021.

MICROSOFT NEWS CENTER BRASIL. **Petrobras começa a usar Inteligência Artificial para aumentar segurança de trabalhadores em operações offshore**. Disponível em: <<https://news.microsoft.com/pt-br/petrobras-comeca-a-usar-inteligencia-artificial-para-aumentar-seguranca-de-trabalhadores-em-operacoes-offshore/>>. Acesso em: 13 set. 2023.

MIGUEL, P. A. C. et al. **Metodologia de pesquisa em engenharia de produção e gestão de operações**. Rio de Janeiro: Elsevier, 2010.

MIRANDA, J. G. V. et al. Maximum angular multiscale entropy: Characterization of the angular self-similarity patterns in two types of SAR images: Oil spills and low-wind conditions images. **Physica D: Nonlinear Phenomena**, v. 455, p. 133892, dez. 2023.
MOSAVI, A. et al. State of the Art of Machine Learning Models in Energy Systems, a Systematic Review. **Energies**, v. 12, n. 7, p. 1301, 4 abr. 2019.

MOUNTRAKIS, G.; IM, J.; OGOLE, C. Support vector machines in remote sensing: A review. **ISPRS Journal of Photogrammetry and Remote Sensing**, v. 66, n. 3, p. 247–259, maio 2011.

MOUSAVIAN, A.; EPPNER, C.; FOX, DI. **6-DOF GraspNet: Variational grasp generation for object manipulation**. Proceedings of the IEEE International Conference on Computer Vision. **Anais...Institute of Electrical and Electronics Engineers Inc.**, 1 out. 2019a.

MOUSAVIAN, A.; EPPNER, C.; FOX, DI. **6-DOF GraspNet: Variational grasp generation for object manipulation**. Proceedings of the IEEE International Conference on Computer Vision. **Anais...Institute of Electrical and Electronics Engineers Inc.**, 1 out. 2019b.

NEGREIROS, A. C. S. V. DE et al. Reliability data analysis of systems in the wear-out phase using a (corrected) q-Exponential likelihood. **Reliability Engineering and System Safety**, v. 197, 1 maio 2020.

NEGREIROS, A. C. S. V. DE et al. Oil spills characteristics, detection, and recovery methods: A systematic risk-based view. **Journal of Loss Prevention in the Process Industries**, v. 80, 1 dez. 2022.

NUNES, C. E. DA S. et al. **MULTICLASIFICAÇÃO DE LOOK-ALIKES EM IMAGENS DE SENSORIAMENTO REMOTO UTILIZANDO REDES NEURAS CONVOLUCIONAIS**. 55º SBPO - LV Simpósio Brasileiro de Pesquisa Operacional, 2023.

OOSTEROM, J.-P.; HALL, C. A. S. Enhancing the evaluation of Energy Investments by supplementing traditional discounted cash flow with Energy Return on Investment analysis. **Energy Policy**, v. 168, p. 112953, set. 2022.

PISNER, D. A.; SCHNYER, D. M. Support vector machine. Em: **Machine Learning**. [s.l.] Elsevier, 2020. p. 101–121.

PLATONOW, V. **Óleo que atingiu praias do Nordeste veio de petroleiro grego, diz PF**. Disponível em: <<https://agenciabrasil.ebc.com.br/geral/noticia/2021-12/oleo-que-atingiu-praias-do-nordeste-veio-de-petroleiro-grego-diz-pf#:~:text=As%20manchas%20de%20%C3%B3leo%20que,localidades%2C%20em%2011%20estados%20litor%C3%A2neos.>>. Acesso em: 10 maio. 2023.

POKHARIYA, H. S.; SINGH, D. P.; PRAKASH, R. Evaluation of different machine learning algorithms for LULC classification in heterogeneous landscape by using remote sensing and GIS techniques. **Engineering Research Express**, v. 5, n. 4, p. 045052, 1 dez. 2023.

RAEISI, A.; AKBARIZADEH, G.; MAHMOUDI, A. Combined Method of an Efficient Cuckoo Search Algorithm and Nonnegative Matrix Factorization of Different Zernike Moment Features for Discrimination between Oil Spills and Lookalikes in SAR Images. **IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing**, v. 11, n. 11, p. 4193–4205, 1 nov. 2018.

RONCI, F. et al. **Oil Spill Detection from SAR Images by Deep Learning**. IGARSS 2020 - 2020 IEEE International Geoscience and Remote Sensing Symposium. **Anais...IEEE**, 26 set. 2020. Disponível em: <<https://ieeexplore.ieee.org/document/9323590/>> Acesso em: 12 maio. 2023.

SCHISTAD SOLBERG, A. H. et al. Automatic detection of oil spills in ERS SAR images. **IEEE Transactions on Geoscience and Remote Sensing**, v. 37, n. 4, p. 1916–1924, 1999.

SELVARAJU, R. R. et al. Grad-CAM: Visual Explanations from Deep Networks via Gradient-Based Localization. **International Journal of Computer Vision**, v. 128, n. 2, p. 336–359, 1 fev. 2020.

SHAKIL, M. H. Environmental, social and governance performance and financial risk: Moderating role of ESG controversies and board gender diversity. **Resources Policy**, v. 72, 1 ago. 2021.

SHARMA, G.; PARASHAR, A.; JOSHI, A. M. DepHNN: A novel hybrid neural network for electroencephalogram (EEG)-based screening of depression. **Biomedical Signal Processing and Control**, v. 66, p. 102393, abr. 2021.

SHEYKHMUSA, M. et al. Support Vector Machine Versus Random Forest for Remote Sensing Image Classification: A Meta-Analysis and Systematic Review. **IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing**, v. 13, p. 6308–6325, 2020.

SIDDIQUI, T.; AMER, A. Y. A. A Comprehensive Review on Text Classification and Text Mining Techniques Using Spam Dataset Detection. Em: **Mathematics and Computer Science Volume 2**. [s.l.] Wiley, 2023. p. 1–17.

SIDEY-GIBBONS, J. A. M.; SIDEY-GIBBONS, C. J. Machine learning in medicine: a practical introduction. **BMC Medical Research Methodology**, v. 19, n. 1, 19 mar. 2019.

SINHA, S.; EBRAHIMI, S.; DARRELL, T. **Variational adversarial active learning**. Proceedings of the IEEE International Conference on Computer Vision. **Anais...**Institute of Electrical and Electronics Engineers Inc., 1 out. 2019.

TOPOUZELIS, K. et al. Detection and discrimination between oil spills and look-alike phenomena through neural networks. **ISPRS Journal of Photogrammetry and Remote Sensing**, v. 62, n. 4, p. 264–270, set. 2007.

UDDIN, S. et al. Comparing different supervised machine learning algorithms for disease prediction. **BMC Medical Informatics and Decision Making**, v. 19, n. 1, 21 dez. 2019.

VERHOEF, P. C. et al. Digital transformation: A multidisciplinary reflection and research agenda. **Journal of Business Research**, v. 122, p. 889–901, jan. 2021.

WANG, C. et al. A labelled ocean SAR imagery dataset of ten geophysical phenomena from Sentinel-1 wave mode. **Geoscience Data Journal**, v. 6, n. 2, p. 105–115, 31 nov. 2019.

WANG, L. et al. Deep generative modeling for mechanistic-based learning and design of metamaterial systems. **Computer Methods in Applied Mechanics and Engineering**, v. 372, p. 113377, dez. 2020a.

WANG, X. et al. CVA²E: A Conditional Variational Autoencoder With an Adversarial Training Process for Hyperspectral Imagery Classification. **IEEE Transactions on Geoscience and Remote Sensing**, v. 58, n. 8, p. 5676–5692, ago. 2020b.

WU, Z. et al. A Comprehensive Survey on Graph Neural Networks. **IEEE Transactions on Neural Networks and Learning Systems**, v. 32, n. 1, p. 4–24, 1 jan. 2021.

XIAN, Y. et al. **F-VAEGAN-D2: A Feature Generating Framework for Any-Shot Learning**. 2019 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR). **Anais...**IEEE, jun. 2019.

YANG, Y. T. et al. The supervised normalized cut method for detecting, classifying, and identifying special nuclear materials. **INFORMS Journal on Computing**, v. 26, n. 1, p. 45–58, dez. 2014.

YAPP, E. K. Y. et al. Comparison of base classifiers for multi-label learning. **Neurocomputing**, v. 394, p. 51–60, 21 jun. 2020.

YEKEEN, S. T.; BALOGUN, A. L. **Advances in remote sensing technology, machine learning and deep learning for marine oil spill detection, prediction and vulnerability assessment**. **Remote Sensing**MDPI AG, , 2 out. 2020.

ZEROUAL, A. et al. Deep learning methods for forecasting COVID-19 time-Series data: A Comparative study. **Chaos, Solitons and Fractals**, v. 140, 1 nov. 2020.

ZHANG, Y.; YEUNG, D. Y. Multilabel relationship learning. **ACM Transactions on Knowledge Discovery from Data**, v. 7, n. 2, 2013.

APÊNDICE A – MATRIZES DE CONFUSÃO MÉTODO 1 (APENAS BANCO DE DADOS PRINCIPAL)

Class 0

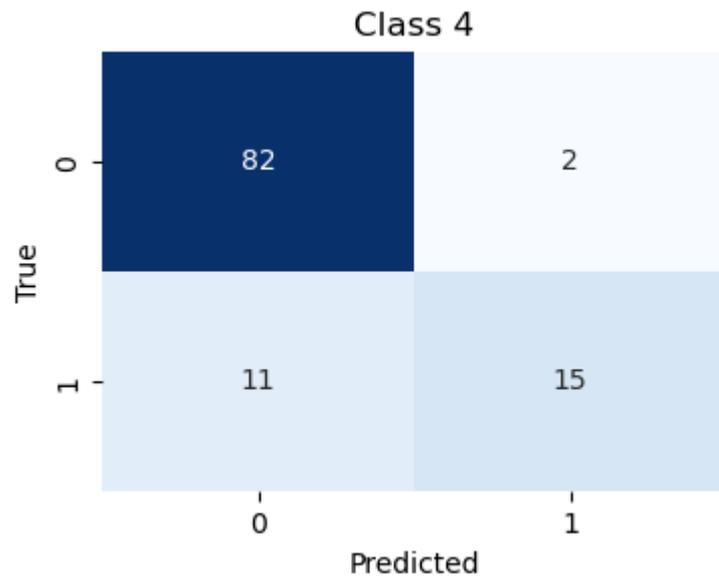
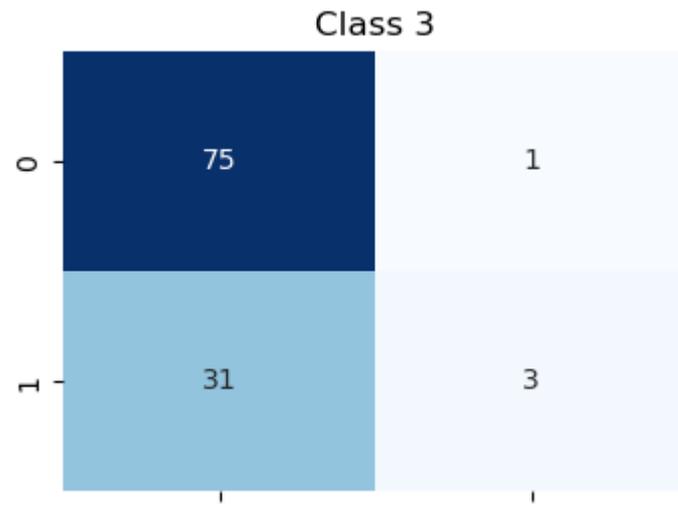
0	0	0
1	0	110

Class 1

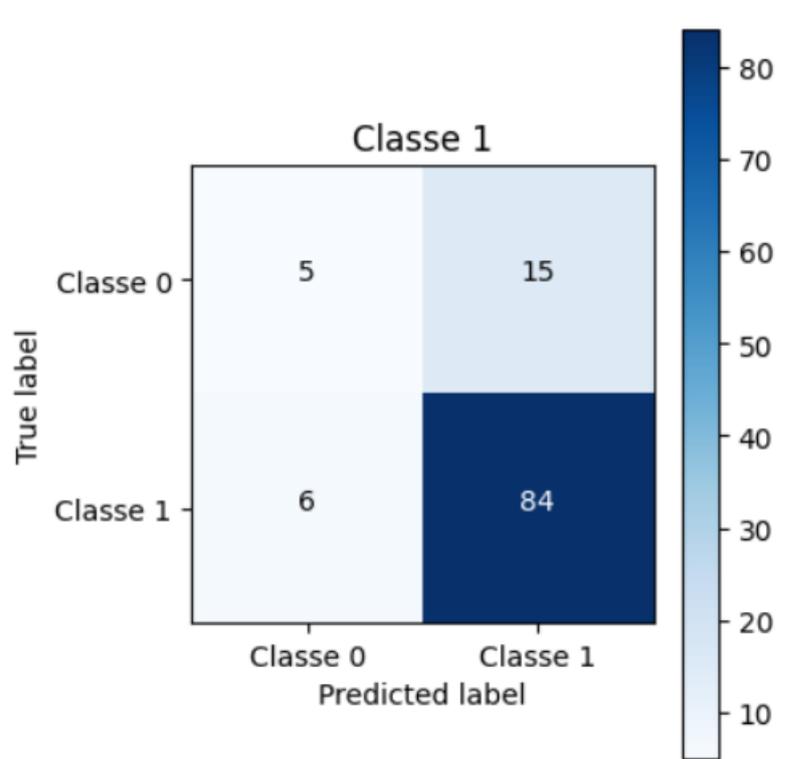
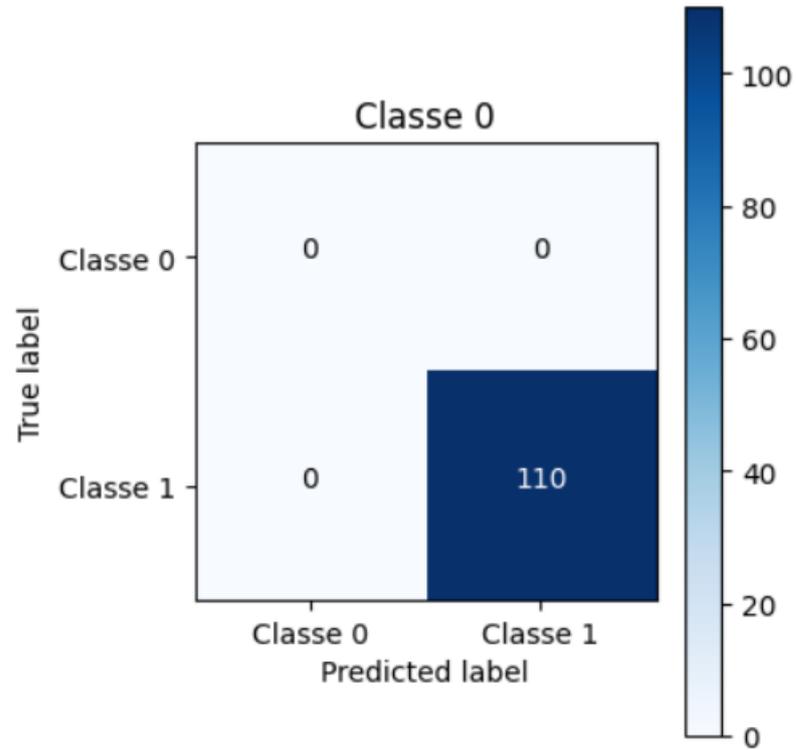
0	5	17
1	3	85

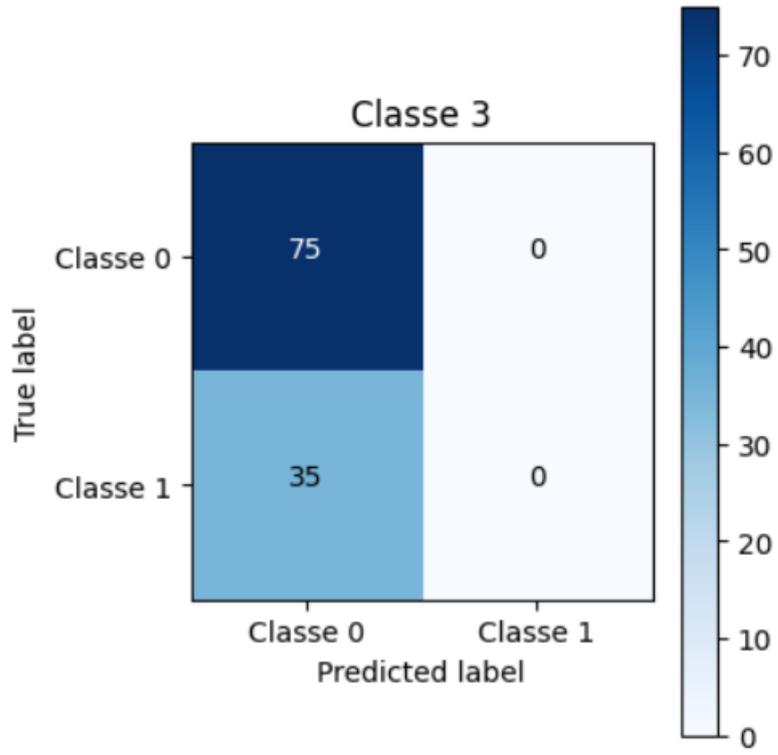
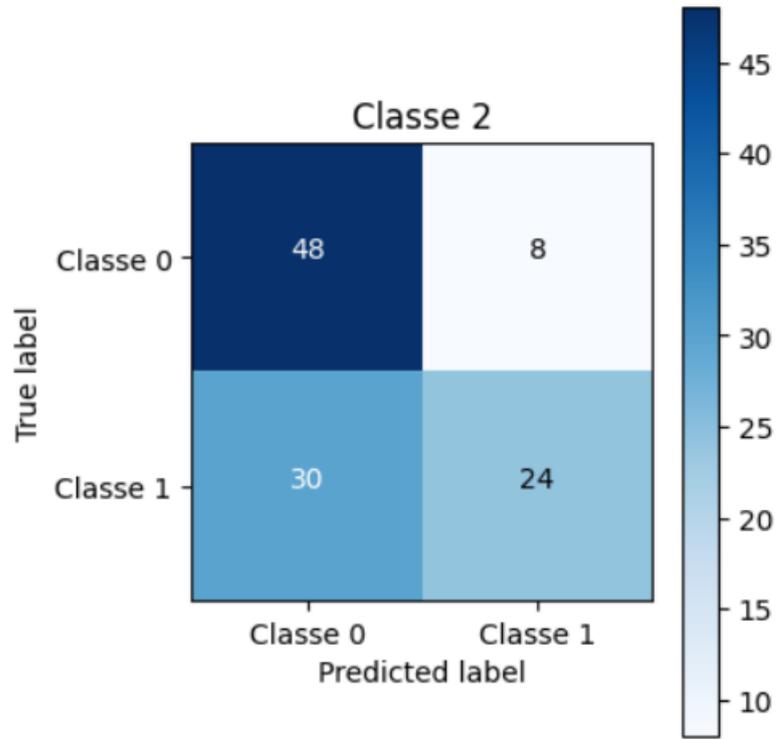
Class 2

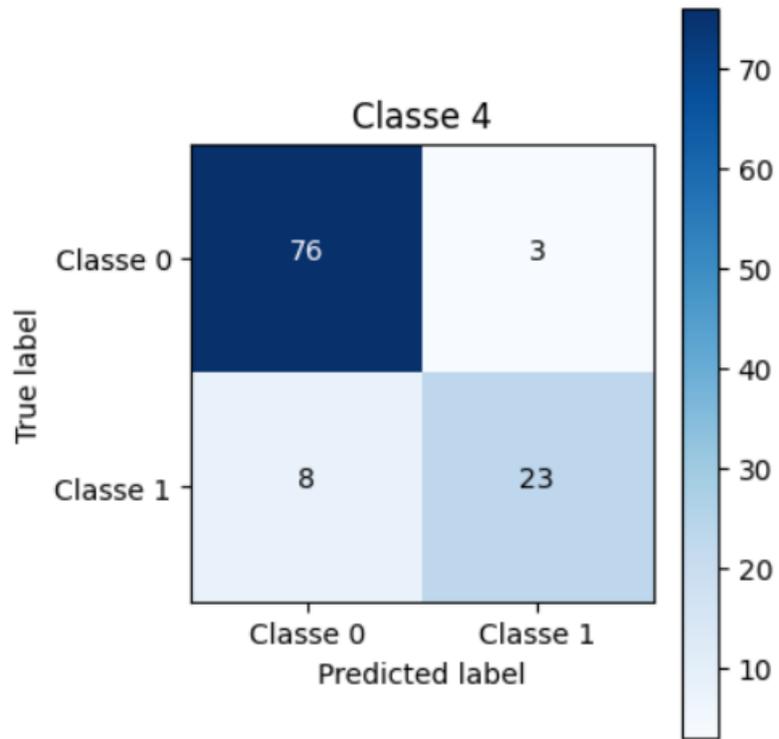
0	47	8
1	31	24



APÊNDICE B – MATRIZES DE CONFUSÃO MÉTODO 2 (APENAS BANCO DE DADOS PRINCIPAL)







APÊNDICE C – MATRIZES DE CONFUSÃO MÉTODO 1 (BANCO DE DADOS PRINCIPAL + BANCO DE DADOS COMPLEMENTAR)

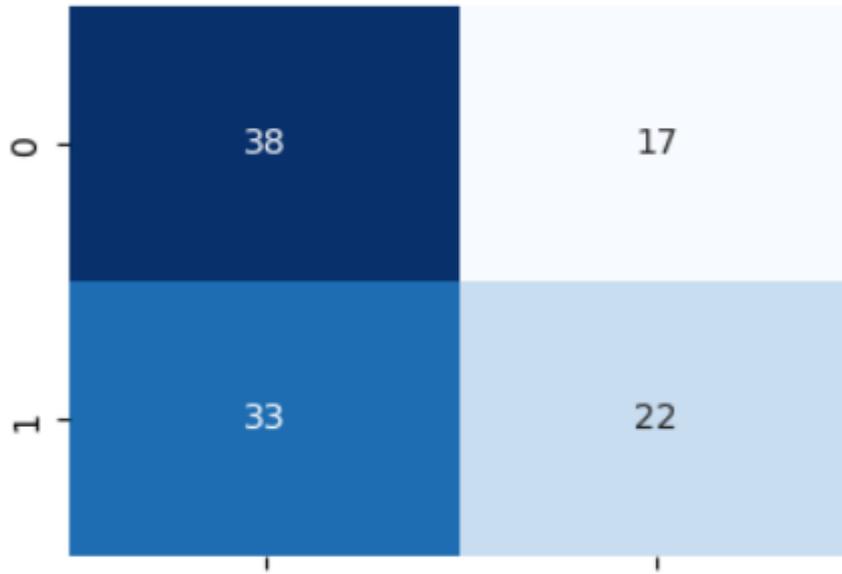
Class 0

0	0	0
1	4	106

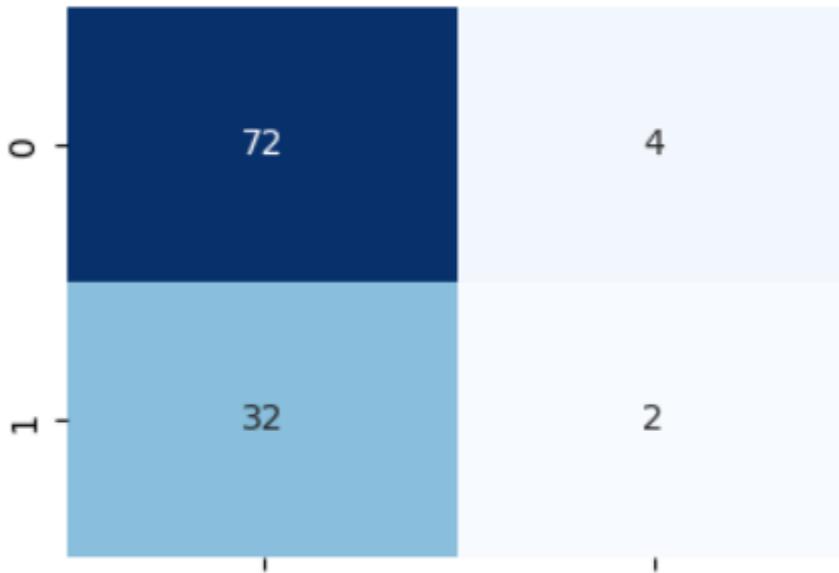
Class 1

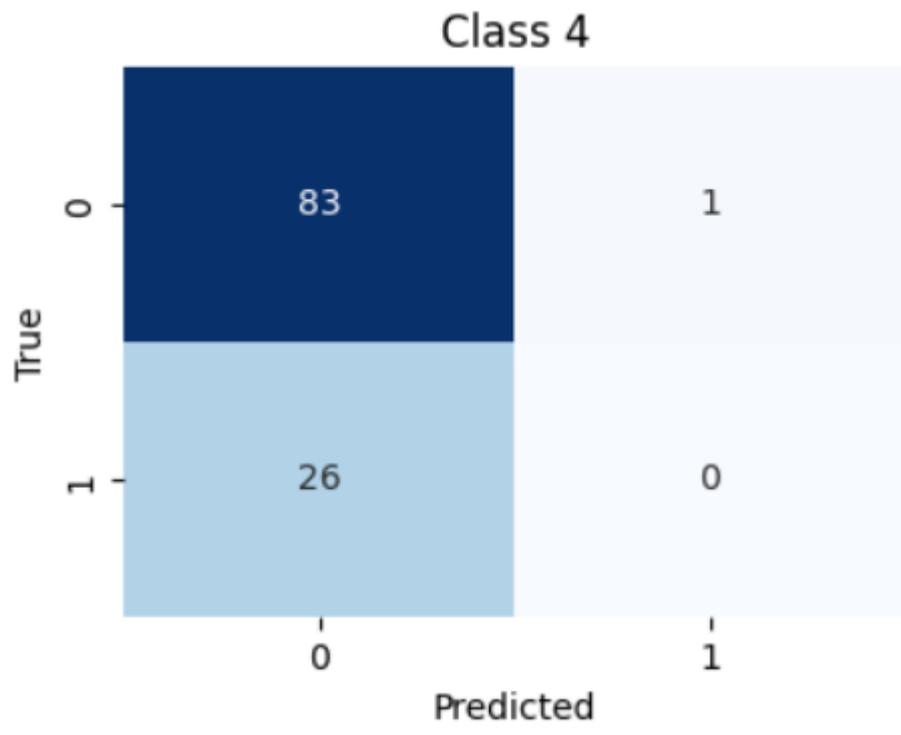
0	4	18
1	2	86

Class 2



Class 3





APÊNDICE D – MATRIZES DE CONFUSÃO MÉTODO 2 (BANCO DE DADOS PRINCIPAL + BANCO DE DADOS COMPLEMENTAR)

