MARCELO SCHNEIDER FIALHO

Aplicação de técnicas de machine learning na análise da correlação entre mudanças climáticas e doenças relacionadas ao clima

São Paulo
2024

MARCELO SCHNEIDER FIALHO

Aplicação de técnicas de machine learning na análise da correlação entre mudanças climáticas e doenças relacionadas ao clima

Monografia apresentada ao Programa de Educação Continuada da Escola Politécnica da Universidade de São Paulo, para obtenção do título de Especialista, pelo Programa de Pós-Graduação em Engenharia de Dados e Big Data.

São Paulo
2024

MARCELO SCHNEIDER FIALHO

Aplicação de técnicas de machine learning na análise da correlação entre mudanças climáticas e doenças relacionadas ao clima

Monografia apresentada ao Programa de Educação Continuada da Escola Politécnica da Universidade de São Paulo, para obtenção do título de Especialista, pelo Programa de Pós-Graduação em Engenharia de Dados e Big Data.

Área de Concentração:

Tecnologia da Informação - Engenharia/ Tecnologia/ Gestão

Orientador:

Luiz Sergio de Souza

São Paulo
2024

Autorizo a reprodução e divulgação total ou parcial deste trabalho, por qualquer meio convencional ou eletrônico, para fins de estudo e pesquisa, desde que citada a fonte.

Este exemplar foi revisado e corrigido em relação à versão original, sob responsabilidade única do autor e com a anuência de seu orientador.

São Paulo, _____ de _____ de _____

Assinatura do autor: _____

Assinatura do orientador: _____

Catalogação-na-publicação

AGRADECIMENTOS

Em primeiro lugar, agradeço a Deus pela força, saúde e sabedoria ao longo desta jornada. Sua presença foi uma constante fonte de inspiração, permitindo-me superar os desafios e seguir em frente. À minha esposa, minha companheira de vida, que esteve ao meu lado em todos os momentos. Obrigado por sua paciência, apoio incondicional e por acreditar em mim, mesmo nos momentos mais difíceis. Sem você, este trabalho não teria sido possível. Agradeço também ao meu orientador, cuja orientação e conhecimento foram fundamentais para o desenvolvimento desta monografia. Sua dedicação e incentivo me motivaram a ir além e a buscar sempre o melhor. Aos meus pais, minha eterna gratidão. Vocês me ensinaram, desde cedo, o valor do estudo e da honestidade. Suas lições foram a base para tudo o que sou e tudo o que conquistei. Obrigado por sempre me incentivarem a ser curioso e íntegro em tudo que faço. A todos que, de alguma forma, contribuíram para a realização deste trabalho, meu sincero agradecimento.

RESUMO

Os avanços tecnológicos nos últimos anos têm possibilitado um grande salto para o poder computacional, permitindo o processamento de grandes bases de dados por meio de algoritmos de Machine Learning (ML) com a finalidade de realizar predições e correlações dentro de um tempo viável. As técnicas de ML têm se destacado como uma ferramenta importante em diversas áreas do conhecimento, promovendo avanços significativos na capacidade de análise e tomada de decisões. Estas técnicas, têm sido aplicadas em campos como a economia, a engenharia, as ciências sociais e a medicina, e se beneficiam da utilização de diferentes algoritmos capazes de aprender e melhorar seu desempenho utilizando grandes bases de dados. No campo da medicina, muitos estudos têm utilizado técnicas de ML para estudar a correlação entre as mudanças climáticas e o surgimento de novas doenças ou surtos de doenças já conhecidas. Entender a relação entre mudanças climáticas e a ocorrência de doenças é de grande importância na implementação de medidas eficazes para a prevenção de surtos e podem contribuir para tomadas de decisões mais assertivas. Este trabalho explorou como as técnicas de ML podem ser utilizadas para identificar correlações entre mudanças climáticas e a ocorrência de doenças relacionadas ao clima, destacando suas implicações para a saúde pública. Foi adotada uma revisão de mapeamento (RM), utilizando a ferramenta Parsifal, para identificar e organizar estudos relevantes na interseção entre ML, mudanças climáticas e saúde. O processo incluiu o planejamento da pesquisa, definição de critérios de inclusão e exclusão, busca sistemática em bases de dados científicas e análise qualitativa dos resultados. Os resultados mostram que modelos como redes neurais artificiais, árvores de decisão e Random Forest têm sido amplamente aplicados com sucesso para prever surtos de doenças como dengue, malária e doenças respiratórias, destacando a influência de variáveis climáticas, como temperatura e padrões de precipitação. Além disso, a análise evidencia que técnicas de ML podem contribuir significativamente para a formulação de políticas de saúde pública mais eficazes, ao proporcionar análises preditivas para a alocação de recursos e estratégias de mitigação. No entanto, limitações relacionadas à qualidade dos dados e à complexidade dos modelos foram identificadas como barreiras a serem superadas. Os achados deste estudo reforçam a importância de integrar abordagens tecnológicas com estratégias interdisciplinares para enfrentar os desafios impostos pelas mudanças climáticas à saúde pública global.

Palavras-Chave – machine learning, mudanças climáticas, doenças.

ABSTRACT

Technological advancements in recent years have enabled a significant leap in computational power, allowing the processing of large datasets using Machine Learning (ML) algorithms to make predictions and correlations within a feasible timeframe. ML techniques have emerged as a vital tool across various fields of knowledge, driving significant progress in analytical capabilities and decision-making processes. These techniques have been applied in fields such as economics, engineering, social sciences, and medicine, benefiting from the use of diverse algorithms capable of learning and improving their performance by leveraging large datasets. In the medical field, numerous studies have employed ML techniques to examine the correlation between climate change and the emergence of new diseases or outbreaks of known diseases. Understanding the relationship between climate change and disease occurrence is of great importance for implementing effective measures to prevent outbreaks and can contribute to more assertive decision-making. This work explored how ML techniques can be used to identify correlations between climate change and the occurrence of climate-related diseases, highlighting their implications for public health. A mapping review (MR) approach was adopted, utilizing the Parsifal tool to identify and organize relevant studies at the intersection of ML, climate change, and health. The process included research planning, defining inclusion and exclusion criteria, conducting systematic searches in scientific databases, and performing qualitative analysis of the results. The findings reveal that models such as artificial neural networks, decision trees, and Random Forests have been widely and successfully applied to predict outbreaks of diseases such as dengue, malaria, and respiratory illnesses, emphasizing the influence of climatic variables such as temperature and precipitation patterns. Furthermore, the analysis demonstrates that ML techniques can significantly contribute to the development of more effective public health policies by providing predictive insights for resource allocation and mitigation strategies. However, limitations related to data quality and the complexity of models were identified as barriers to be addressed. The findings of this study underscore the importance of integrating technological approaches with interdisciplinary strategies to tackle the challenges posed by climate change to global public health.

Keywords – machine learning, climate change, diseases.

LISTA DE FIGURAS

1 Classificação de algoritmos de ML. 16

2 Representação da linha de regressão linear 19

3 Representação da linha de regressão logística 21

4 Representação de um perceptron. 22

5 Representação de uma rede neural. 23

6 Separação do tema em palavras-chave com método PICOC 45

7 Gerenciamento das palavras-chave e sinônimos no Parsifal 45

8 Criação da string de busca por meio do Parsifal 48

9 Critérios de Inclusão e Exclusão na ferramenta Parsifal 49

10 Seleção de estudos na ferramenta Parsifal. 51

11 Processo de seleção dos trabalhos. 52

12 Número de trabalhos por ano 53

13 Tipos de publicação 53

14 Utilização dos algoritmos de acordo com o campo de estudo 58

15 Utilização de cada algoritmo. 59

LISTA DE TABELAS

1 Comparação entre tipos de revisão. 37

2 Palavras-chave e sinônimos 46

3 Avaliação de aderência 46

4 Adaptação da string de busca para cada base de dados. 48

5 Critérios de Inclusão e Exclusão 49

6 Resultado da busca utilizando a string de busca SB1. 50

7 Trabalhos selecionados após a leitura do título e resumo. 52

8 Número de trabalhos selecionados em cada etapa de seleção. 52

9 Número de artigos por autor. 54

10 Estudos selecionados e seus respectivos autores. 55

11 Relação entre estudos selecionados e algoritmos mais utilizados. 59

SUMÁRIO

1 Introdução **10**

1.1 Motivação 11

1.2 Objetivo 11

1.2.1 Objetivos específicos 11

1.3 Justificativa 12

1.4 Metodologia 12

2 Fundamentação Teórica **15**

2.1 Machine Learning 15

2.1.1 Regressão Linear 18

2.1.2 Regressão Logistica 20

2.1.3 Redes Neurais Artificiais 21

2.1.4 K-Means 23

2.1.5 Árvores de decisão 25

2.1.6 Random Forest 26

2.2 Mudanças climáticas 27

2.3 Doenças relacionadas ao clima 29

2.3.1 Doenças transmitidas por vetores 30

2.3.2 Doenças de origem hídrica 30

2.3.3 Doenças respiratórias 31

2.3.4 Doenças Cardiovasculares e relacionadas ao calor 31

2.3.5 Saúde mental e impactos psicológicos 32

2.4 Sistemas de saúde e sua atuação no tratamento de doenças 33

2.5 Revisão de Mapeamento 36

2.5.1 Parsifal 40

2.5.1.1 Planejamento 41

2.5.1.2 Condução 42

2.5.1.3 Documentação 42

3 Processo de Revisão de Mapeamento **44**

3.1 Planejamento da Revisão de Mapeamento (RM) 44

3.2 Condução 50

4 Resultados **53**

4.1 Sumarização dos resultados 53

4.2 Algoritmos 57

4.3 Capacidade de detecção e previsão 60

4.4 Políticas de saúde pública 61

5 Conclusão **62**

Referências **64**

1 INTRODUÇÃO

De acordo com Calvin et al. (2023), o aumento das temperaturas globais, as alterações nos padrões de precipitação e a intensificação de eventos climáticos extremos têm demonstrado uma relação direta com a propagação de diversas doenças. Essa interação complexa entre clima e saúde é um campo de pesquisa em crescimento, no qual o uso de Machine Learning (ML) se destaca como uma importante ferramenta.

Doenças relacionadas ao clima, como as doenças respiratórias, cardiovasculares e as infecciosas transmitidas por vetores, têm mostrado um aumento significativo em sua incidência, especialmente em regiões vulneráveis. O estudo realizado por Ebi e McGregor (2009) destaca que a exposição a temperaturas extremas pode agravar doenças crônicas e aumentar a mortalidade, enquanto a modificação de ecossistemas resulta em mudanças na distribuição de vetores, potencializando a propagação de doenças infecciosas. Nesse contexto, o uso de ML permite a análise de grandes volumes de dados climáticos e epidemiológicos, possibilitando a identificação de padrões que não seriam facilmente perceptíveis através de métodos estatísticos tradicionais.

A literatura aponta que técnicas de ML, como redes neurais, árvores de decisão e modelos de regressão, têm sido aplicadas com sucesso para prever surtos de doenças e entender suas relações com variáveis climáticas (RACLOZ et al., 2012; FAHIM et al., 2022a). No entanto, apesar do potencial promissor, a implementação de ML na análise de correlações entre mudanças climáticas e doenças ainda enfrenta desafios, como a qualidade dos dados, a complexidade dos modelos e a necessidade de interpretações confiáveis para a tomada de decisão em saúde pública (WANG et al., 2023).

Diante desse cenário, este trabalho tem como objetivo realizar uma revisão de mapeamento da literatura existente sobre a utilização de ML para identificar a correlação entre mudanças climáticas e doenças ligadas ao clima. A revisão busca mapear as principais abordagens, resultados e lacunas no conhecimento, fornecendo uma base para futuras pesquisas e aplicações práticas no enfrentamento dos desafios impostos pelas mudanças climáticas e à saúde pública.

1.1 Motivação

A relação entre mudanças climáticas e a ocorrência de doenças relacionadas ao clima é uma área de pesquisa de crescente importância, devido ao impacto significativo que as alterações climáticas têm sobre a saúde pública global. Estudos indicam que variáveis climáticas, como temperatura, umidade e padrões de precipitação, influenciam diretamente a ocorrência e a distribuição de doenças como malária, dengue e doenças respiratórias. Esse cenário destaca a necessidade de compreender essas interações para prever surtos e implementar medidas preventivas eficazes. Essas técnicas permitem a análise de grandes volumes de dados climáticos e de saúde, identificando padrões complexos que não seriam perceptíveis através de métodos estatísticos tradicionais. A literatura aponta que algoritmos de aprendizado de máquina, como redes neurais, árvores de decisão e modelos de regressão, têm sido aplicados com sucesso para prever surtos de doenças e entender suas relações com variáveis climáticas (RACLOZ et al., 2012; FAHIM et al., 2022a). No entanto, apesar do potencial promissor, a implementação dessas técnicas ainda enfrenta desafios, como a qualidade dos dados, a complexidade dos modelos e a necessidade de interpretações confiáveis para a tomada de decisão na área da saúde pública (WANG et al., 2023).

1.2 Objetivo

Este trabalho tem como objetivo principal a realização de uma Revisão de Mapeamento (RM) com o intuito de explorar a interseção entre mudanças climáticas e a incidência de doenças através do uso de técnicas de Machine Learning (ML). Especificamente, o trabalho buscará os seguintes objetivos específicos:

1.2.1 Objetivos específicos

1. **Algoritmos:** Explorar e identificar os algoritmos e modelos de ML utilizados para analisar grandes conjuntos de dados climáticos e de saúde, compreendendo a capacidade destes algoritmos de captar padrões complexos e gerar previsões acuradas.

2. **Capacidade de Detecção e Previsão:** Examinar a capacidade das técnicas de ML em detectar correlações significativas entre as variabilidades climáticas - como temperatura, umidade e padrões de precipitação - e a incidência de doenças sensíveis ao clima, como malária, dengue e doenças respiratórias.

3. **Políticas de Saúde Pública:** Entender como os resultados obtidos através das análises com ML podem ser utilizados para informar e aprimorar políticas de saúde pública e estratégias de mitigação, considerando a qualidade das interpretações geradas pelos modelos, a qualidade dos dados utilizados, e a aplicabilidade prática dos insights para a tomada de decisão em contextos de saúde pública.

1.3 Justificativa

As mudanças climáticas impactam significativamente na saúde pública, afetando a incidência e distribuição de várias doenças como malária, dengue e doenças respiratórias. Estas doenças são particularmente sensíveis às variabilidades climáticas, incluindo temperatura, umidade e padrões de precipitação. Assim, a previsão de surtos e a implementação de medidas preventivas eficazes demandam uma compreensão aprofundada das interações entre clima e saúde.

Dentro desse cenário, as técnicas de ML emergem como ferramentas poderosas capazes de analisar grandes volumes de dados climáticos e de saúde, permitindo a identificação de padrões complexos e a geração de previsões acuradas. Algoritmos como redes neurais, árvores de decisão e modelos de regressão já demonstraram sucesso na previsão de surtos de doenças e na análise de suas correlações com variáveis climáticas. Contudo, apesar de seu potencial, a implementação dessas técnicas enfrenta desafios significativos, tais como a qualidade dos dados disponíveis, a complexidade dos modelos e a necessidade de interpretações confiáveis para a tomada de decisão em saúde pública. É precisamente devido a esses desafios e ao vasto corpo de literatura existente que a realização de uma Revisão de Mapeamento (RM) se torna necessária. Permitindo assim, consolidar o conhecimento atual, identificar as lacunas existentes e fornecer uma visão abrangente das principais abordagens e resultados obtidos até o momento.

1.4 Metodologia

A Revisão de Mapeamento (RM) tem como objetivo o mapeamento da literatura existente sobre um tema, identificando lacunas, tendências de pesquisa e tópicos que ainda precisam ser explorados. Este tipo de revisão é menos detalhada do que uma revisão sistemática da literatura (RSL), porém, fornece uma visão geral do campo de estudo, sendo útil em áreas onde novas tecnologias estão emergindo rapidamente, como Engenharia de Software, Nanotecnologia, Robótica e Inteligência Artificial. Revisões de mapeamento são

amplamente utilizadas para identificar quais técnicas estão sendo aplicadas em diferentes áreas.

O processo delineado por Kitchenham (2004) pode ser adaptado para uma RM, embora o objetivo de uma RM seja mais geral, propiciando o mapeamento do estado da arte e identificando lacunas e tendências de pesquisa, em vez de responder perguntas de pesquisa específicas e detalhadas. Nesta RM, a ferramenta Parsifal é utilizada para auxiliar na condução da revisão de mapeamento.

As principais etapas de uma RM a serem seguidas por este trabalho são delineadas abaixo:

1. **Planejamento:**
   - **Definição do Objetivo e Escopo:** No processo de Kitchenham (2004), a fase de planejamento é fundamental para garantir que a revisão siga uma abordagem sistemática. Em uma revisão de mapeamento, o objetivo é explorar um campo amplo de pesquisa em vez de responder a questões específicas. Desta forma, o planejamento deve definir claramente as áreas de investigação e as categorias a serem mapeadas, como tecnologias emergentes, métodos ou temas ainda não explorados.
   - **Critérios de Inclusão/Exclusão:** Diferente de uma RSL, onde os critérios são mais restritivos para garantir a resposta a uma questão específica, no mapeamento os critérios podem ser mais amplos, buscando capturar uma maior diversidade de estudos para fornecer uma visão geral. Os critérios devem ser definidos de modo a incluir estudos relevantes para o objetivo e escopo estabelecidos.
2. **Condução da Revisão:**
   - **Busca Sistemática:** Utilização de bases de dados acadêmicas e científicas para identificar estudos relevantes. A ferramenta Parsifal será utilizada para gerenciar o processo de busca, seleção e extração dos dados. A busca deve ser abrangente, utilizando termos e palavras-chave relacionados ao tema da pesquisa.
   - **Seleção de Estudos:** Aplicação dos critérios de inclusão e exclusão previamente definidos para selecionar os estudos mais relevantes. Esta etapa envolve a triagem inicial dos títulos e resumos, seguida pela leitura completa dos artigos selecionados.

- **Extração de Dados:** Coleta de informações relevantes dos estudos incluídos na revisão. Esta etapa possui como objetivo, mapear as tendências de pesquisa e identificar lacunas no conhecimento atual.

3. **Análise e Síntese:**

   - **Análise dos Dados:** Análise qualitativa e quantitativa dos dados extraídos, identificando padrões, tendências e lacunas na literatura. Ferramentas estatísticas e de visualização de dados podem ser utilizadas para auxiliar na análise.
   - **Síntese dos Resultados:** Integração dos achados da revisão em um formato coerente e compreensível, destacando as principais contribuições dos estudos analisados, bem como as lacunas e áreas para futuras pesquisas.

4. **Relato dos Resultados:**

   - **Elaboração do Relatório:** Preparação de um relatório detalhado dos achados da revisão, incluindo a descrição do processo metodológico, resultados da análise e síntese dos dados, e discussão das implicações para a pesquisa e prática.
   - **Disseminação dos Resultados:** Divulgação dos achados da revisão em conferências, revistas acadêmicas e outros meios relevantes para a comunidade científica.

2 FUNDAMENTAÇÃO TEÓRICA

Neste capítulo são abordados alguns conceitos necessários para o desenvolvimento da metodologia proposta por este trabalho. A sessão 2.1, apresenta uma introdução sobre Machine Learning e sua aplicação. Na sessão 2.2 são abordados conceitos relacionados a mudanças climáticas e seus efeitos. Já na sessão 2.4, é apresentada uma introdução a respeito do sistema de saúde e doenças. Por fim, a seção 2.5 aborda conceitos sobre a Revisão de Mapeamento, bem como a ferramenta Parsifal, utilizada como apoio no processo da revisão de mapeamento realizada neste trabalho.

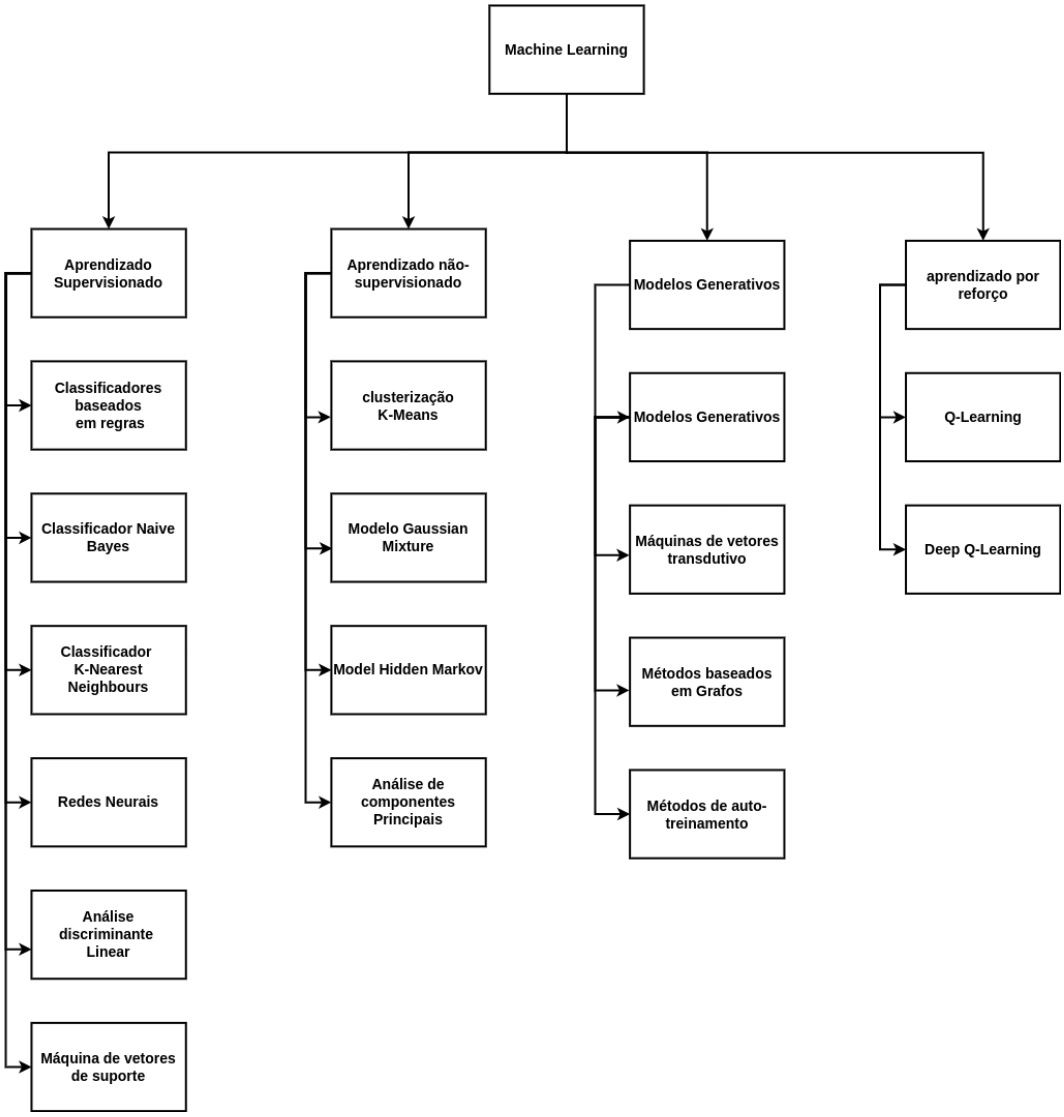
2.1 Machine Learning

O rápido aumento na quantidade de dados gerados e coletados tem impulsionado a necessidade de desenvolver abordagens eficientes para sua análise. Nesse contexto, Machine Learning (ML) tem se destacado como uma área da ciência da computação dedicada ao desenvolvimento de algoritmos capazes de permitir que máquinas aprendam padrões a partir de dados - sem programação explícita - para tarefas específicas. Desta forma, ML tem sido amplamente aplicada em diferentes áreas, como reconhecimento de padrões, processamento de linguagem natural e previsão (N; GUPTA, 2020).

Além disso, segundo N e Gupta (2020), a sua ampla aplicabilidade promoveu diferentes abordagens que se diferenciam pela forma como o aprendizado do algoritmo é realizado para a resolução de um determinado problema. Tais abordagens podem ser observadas na figura 1 e são classificadas de acordo com o tipo de aprendizado da seguinte forma:

- **Aprendizado supervisionado:** neste modo de aprendizagem, é apresentado ao modelo de ML um conjunto de dados para as variáveis de entrada e outro conjunto de dados rotulados para as variáveis de saída durante o processo de treinamento. Desta forma, espera-se que o modelo de ML realize o mapeamento entre os dados de entrada e saída, de forma que o modelo seja capaz de generalizar os valores de saída para novas entradas desconhecidas. Existem diferentes tipos de algoritmos que

Figura 1: Classificação de algoritmos de ML.



Fonte: Adaptado de N e Gupta (2020).

utilizam esta técnica de aprendizado, e alguns deles estão representados na Figura 1.

- **Aprendizado não-supervisionado:** diferente do aprendizado supervisionado, no aprendizado não supervisionado os dados utilizados não possuem rótulos ou classificação pré-definidas. Neste caso, o objetivo do modelo é classificar os dados, criando clusters, com base em sua similaridade. Na Figura 1, é possível observar diferentes tipos de algoritmos que utilizam este tipo de aprendizado

- **Aprendizado semi-supervisionado:** segundo Engelen e Hoos (2020) o aprendizado semi-supervisionado utiliza não apenas dados rotulados como também dados não-rotulados em seu processo de aprendizagem. Nesta abordagem, a premissa é

que a utilização de uma grande quantidade de dados não rotulados, juntamente com uma pequena quantidade de dados rotulados, seja capaz de melhorar o desempenho do modelo de aprendizagem. Um exemplo da utilização desta abordagem é a mineração de texto, onde na maioria das vezes há uma pequena quantidade de textos rotulados e um grande volume de documentos rotulados. Exemplos de algoritmos que utilizam esta abordagem pode ser visto na Figura 1

- **Aprendizado por reforço:** no aprendizado por reforço, um agente aprende a realizar ações em um ambiente, recebendo feedback na forma de recompensas ou punições após cada ação. Neste caso, o objetivo do agente é descobrir uma política que maximize a recompensa acumulada ao longo do tempo. Segundo N e Gupta (2020) o aprendizado por reforço possui aplicações em diferentes áreas como robótica, onde os agentes podem aprender a navegar e interagir em ambientes físicos, assim como em jogos, para treinar agentes capazes de superar jogadores humanos. Na Figura 1, são apresentados alguns exemplos de algoritmos utilizados no aprendizado por reforço.

Como visto anteriormente, diferentes técnicas de ML tem sido utilizadas objetivando a solução de problemas específicos. Estas diferentes abordagens tem proporcionado a sua aplicação em diferentes áreas do conhecimento, é o caso de Campbell et al. (2020), onde é utilizado um modelo de aprendizado de máquina baseado em Random Forest, com o objetivo de explorar o uso de variáveis climáticas derivadas de satélites para prever o risco ambiental de surtos de cólera na costa da Índia. Para tanto, os autores analisaram dados de surtos de cólera de 2010 a 2018 e descobriram que variáveis como concentração de clorofila-a, salinidade da superfície do mar e temperatura da superfície terrestre são fortes preditores de surtos. Além disso, o modelo alcançou uma precisão de 0,99 e uma pontuação F1 de 0,942, mostrando-se altamente eficaz na previsão de surtos, com 89,5% de sensibilidade. Os resultados do trabalho sugerem ainda que o uso de técnicas de ML pode melhorar significativamente os sistemas de monitoramento de risco de surtos de cólera, especialmente em áreas costeiras. Contudo, os autores apontam a necessidade de mais estudos que incorporem dados socioeconômicos e climáticos para aumentar a robustez e aplicabilidade do modelo em diferentes regiões e sob diferentes condições climáticas.

Já em Berrang-Ford et al. (2021), foi realizada uma análise sistemática da vasta literatura científica a respeito da relação entre mudanças climáticas e saúde. Segundo os autores, dada a rápida expansão desse campo de pesquisa, os métodos tradicionais de revisão sistemática tornaram-se insuficientes. Desta forma, foram aplicadas técnicas

de aprendizado supervisionado e não supervisionado, com a finalidade de classificar as publicações em três principais temas. Para tanto, os autores utilizaram algoritmos de aprendizado supervisionado para treinar um modelo que identificasse quais artigos eram relevantes para os objetivos do estudo. Assim, um conjunto de dados inicial foi rotulado manualmente por especialistas, categorizando os estudos em três grandes áreas: impactos climáticos, adaptação e mitigação. Esse conjunto rotulado serviu como base de treinamento para o algoritmo, que então previu a relevância dos artigos restantes de forma automatizada. Já a utilização de aprendizagem não supervisionada, foi aplicada para identificar padrões e agrupamentos temáticos na literatura. Isso permitiu a identificação automática de clusters de palavras relacionadas, que revelaram os principais temas de pesquisa emergentes na interseção entre mudanças climáticas e saúde, como eventos climáticos extremos, qualidade do ar e doenças infecciosas. Segundo os autores, a aplicação dessas técnicas não apenas acelerou o processo de análise, mas também aumentou a precisão e a abrangência da revisão, proporcionando uma visão mais detalhada e completa da relação entre mudanças climáticas e saúde.

Um outro exemplo, porém utilizando a técnica de aprendizagem por reforço, pode ser encontrado em Razdan e Mittal (2024), que procura explorar os impactos de mudanças climáticas nos campos de colheita de soja e milho, empregando técnicas de Machine Learning. Para tanto os autores utilizaram um conjunto de dados composto por 28242 registros obtidos entre os anos de 1990 e 2013, dos quais 22593 amostras foram utilizadas para treino e 5649 amostras para os testes. Estes dados foram utilizados no treinamento de diferentes modelos, entre os quais, o treinamento por reforço utilizando o algoritmo Random forest que obteve a melhor performance, atingindo um R^2 com valor de 0.97, demonstrando uma boa taxa de eficiência.

Os tipos de aprendizado mencionados pelos trabalhos anteriores podem ser aplicados por meio de diferentes tipos de algoritmos. A seções seguintes fazem uma breve descrição destes algoritmos e o seu funcionamento.

2.1.1 Regressão Linear

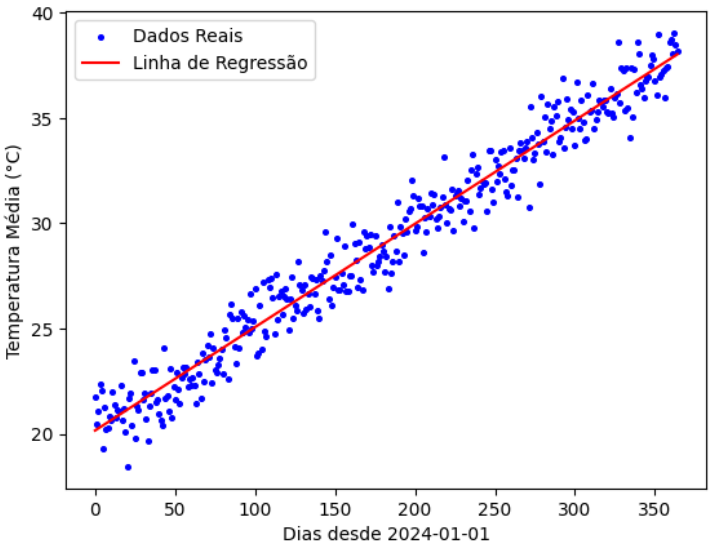
A regressão linear é um dos algoritmos mais simples e amplamente utilizados para problemas de previsão contínua. Esse algoritmo busca modelar a relação entre uma variável dependente e uma ou mais variáveis independentes, minimizando o erro quadrático médio. Este algoritmo é adequado para problemas onde a relação entre as variáveis é aproximadamente linear (CHATTERJEE; HADI, 1986). Além disso, a regressão linear

é utilizada para prever valores futuros com base em dados históricos, entender a relação entre variáveis e identificar fatores que influenciam em um resultado específico. Por exemplo, em marketing, a regressão linear pode ser utilizada para prever as vendas com base em gastos com publicidade e outros fatores (SKIERA et al., 2022).

$$y = \beta_0 + \beta_1^x + \epsilon \tag{2.1}$$

Desta forma, a regressão linear busca modelar a relação entre uma variável dependente y e uma ou mais variáveis independentes x, assumindo uma relação linear entre elas. A equação básica da regressão linear simples é representada pela expressão 2.1, onde: y é a variável dependente, ou seja, o que se quer prever; x é a variável independente; β_0 é o intercepto; β_1 é o coeficiente de inclinação (taxa de variação); e ϵ é o termo de erro.

Figura 2: Representação da linha de regressão linear



Fonte: Autor.

Desta forma, o modelo é ajustado minimizando o erro quadrático médio, ou seja, a diferença entre os valores observados e os valores preditos. A Figura 2, utiliza dados fictícios para representar uma regressão linear que correlaciona a temperatura média global e a contagem de dias desde 1 de Janeiro de 2024. Neste gráfico, a reta de regressão em vermelho é a reta que melhor se ajusta ao conjunto de dados, representando uma função por meio da qual é possível prever valores para outros conjunto de dados. Neste caso a função para a linha de regressão linear é representada por $y = 0.491x + 20.16$, com $R^2 = 0.965$, onde R^2 é uma métrica utilizada para avaliar a qualidade do ajuste de um

modelo de regressão, indicando a proporção da variância na variável dependente que é previsível a partir da variável independente.

2.1.2 Regressão Logistica

A Regressão Logística é uma técnica estatística utilizada para classificação binária, convertendo uma função linear em uma probabilidade limitada entre 0 e 1 por meio da função sigmoide. Esta técnica de ML possui a finalidade de modelar a relação entre uma variável dependente binária e uma ou mais variáveis independentes. (WILSON; LORENZ, 2015). Ao contrário da regressão linear, que prevê valores contínuos, a regressão logística possui como finalidade prever a probabilidade de um evento binário.

$$p = \frac{1}{1 + e^{-z}} \tag{2.2}$$

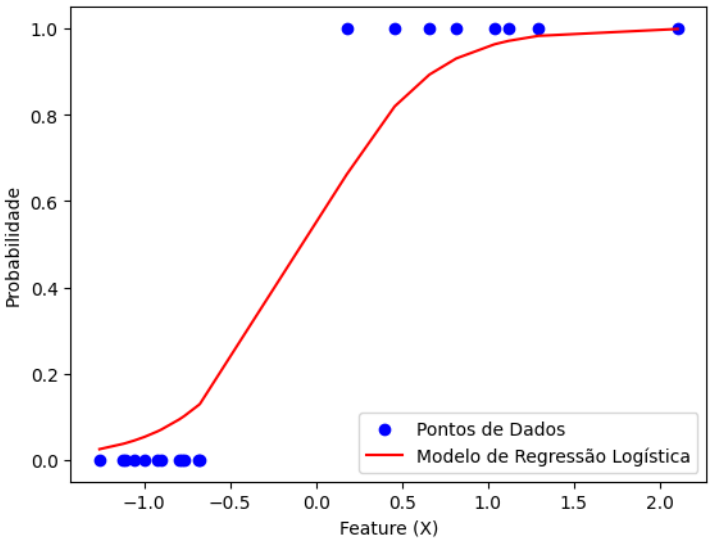
Dentro do contexto de ML, a regressão logística é representada pela expressão, 2.2, onde: p é a probabilidade do evento ocorrer; e é a constante de Euler (aproximadamente 2.71828); e z é uma combinação linear dos preditores (variáveis independentes) e seus coeficientes. Esta função sigmoide transforma os valores previstos em probabilidades, o que é particularmente útil para classificação binária em ML.

Desta forma, a utilização da regressão logística é indicada em casos onde:

- as variáveis de resposta são binárias, podendo assumir apenas dois valores como sim/não ou sucesso/fracasso. Um exemplo de sua utilização é a previsão da probabilidade de um cliente fazer uma compra (WILSON; LORENZ, 2015);
- o objetivo é prever a probabilidade de um determinado evento ocorrer (WILSON; LORENZ, 2015);
- as respostas não são distribuídas de maneira contínua, como por exemplo, prever se um paciente terá uma doença com base em várias características (WILSON; LORENZ, 2015).

A Figura 3, representa o ajuste de um modelo de regressão logística a um conjunto de dados binários, onde os pontos azuis representam o conjunto de dados de teste podendo assumir um valor de 0 a 1. Já no eixo x, estão as variáveis independentes, utilizadas para

Figura 3: Representação da linha de regressão logística



Fonte: Autor.

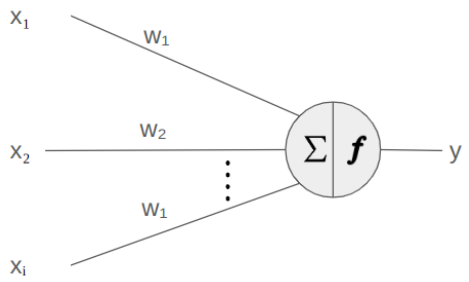
prever a classe, enquanto o eixo y é representado pelos valores observados para a variável dependente binária sendo 0 ou 1. A linha vermelha representa as probabilidades preditas pelo modelo de regressão logística. Esta linha é uma curva sigmoide, característica da função logística, este formato sigmoide é importante, pois a regressão logística transforma uma função linear em uma curva que varia suavemente entre 0 e 1, o que corresponde a uma probabilidade. Desta forma, quanto mais à direita o ponto estiver, maior a probabilidade de ser classificado como 1 e quanto mais à esquerda, maior a chance de ser 0.

2.1.3 Redes Neurais Artificiais

O conceito de redes neurais artificiais (RNA) foi formalizado pela primeira vez por McCulloch e Pitts (1943), onde foi desenvolvido um modelo matemático simplificado de um neurônio que recebia múltiplas entradas, aplicava uma função de ativação (geralmente um somatório simples), e produzia uma saída binária. Este é um modelo de processamento básico que tenta simular como os neurônios biológicos disparam quando recebem certos estímulos, sendo considerado o precursor das redes neurais e a base para o desenvolvimento posterior.

Na década de 1950, o pesquisador Frank Rosenblatt introduziu o Perceptron, um modelo simples de rede neural que podia aprender a partir de exemplos. O perceptron consiste em um único neurônio capaz de classificar dados linearmente separáveis, ajustando seus pesos de entrada por meio de um processo iterativo de aprendizado, conforme

Figura 4: Representação de um perceptron.



Fonte: Autor.

representado na Figura 4, onde, x_i são as entradas, w_i são os pesos aplicados para cada entrada e y é a saída (RUMELHART et al., 1986). No entanto, devido à incapacidade do perceptron de resolver problemas não lineares, como o problema do XOR, o entusiasmo pelas RNAs diminuiu na década de 1970, após o trabalho crítico de Minsky e Papert (1969).

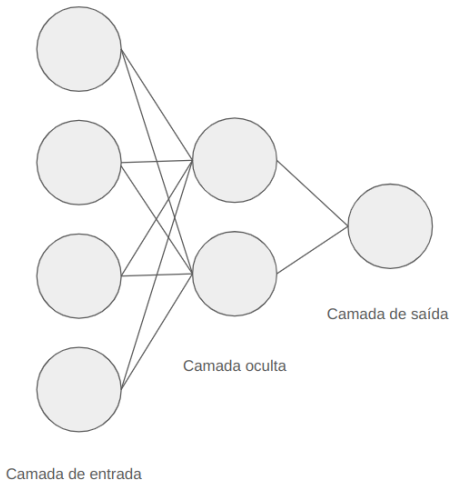
Somente em 1986, com a introdução do algoritmo de backpropagation, por Rumelhart, Hinton e Williams, as redes neurais ressurgiram como uma ferramenta importante. Este algoritmo permite que redes multicamadas ajustem seus parâmetros de maneira eficiente, solucionando problemas não lineares e permitindo que as RNAs evoluíssem para formas mais complexas e profundas (RUMELHART et al., 1986).

As RNAs são formadas por unidades interconectadas, como pode ser visto na Figura 5, estas unidades são chamadas neurônios artificiais, e são organizadas em três tipos de camadas principais:

- camada de entrada: Recebe os dados de entrada;
- camadas ocultas: processam as informações, ajustando os pesos das conexões por meio de um algoritmo de aprendizado. Essas camadas intermediárias são responsáveis por capturar as relações complexas entre os dados.
- camadas de saída: produz a previsão final ou classificação com base nas entradas processadas pelas camadas ocultas.

Cada neurônio realiza uma combinação linear das entradas recebidas, multiplicando os valores das entradas pelos pesos correspondentes e somando-os. A saída resultante passa por uma função de ativação, como a função sigmoide, ReLU (Rectified Linear Unit), ou a tangente hiperbólica, que introduzem não-linearidade ao modelo, permitindo que as RNAs

Figura 5: Representação de uma rede neural.



Fonte: Autor.

aprendam relações complexas. Desta forma, a saída de um neurônio y pode ser descrita pela equação 2.3, onde: x_i são as entradas; w_i são os pesos associados às entradas; b é o viés; f é a função de ativação; e y é a saída do neurônio.

$$y = f\left(\sum_{i=1}^{n} w_i x_i + b\right) \tag{2.3}$$

O aprendizado de uma RNA ocorre ajustando os pesos w_i e o viés b por meio de um processo chamado treinamento supervisionado, onde os pesos são ajustados para minimizar o erro entre as saídas previstas e os valores reais dos dados de treinamento. Além disso, o algoritmo de backpropagation, juntamente com o método de gradiente descendente, é utilizado para ajustar iterativamente esses pesos, reduzindo o erro global da rede.

Devido a sua versatilidade, as RNAs tem sido aplicadas em várias áreas como reconhecimento de padrões em imagens, processamento de linguagem natural, diagnóstico médico e previsão financeira (LECUN et al., 2015; CUI et al., 2023; PATEL et al., 2014; HOCHREITER; SCHMIDHUBER, 1997)

2.1.4 K-Means

O algoritmo K-means é um dos métodos mais utilizados em ML e mineração de dados para agrupamento. Este algoritmo utiliza uma forma de aprendizado não supervisionado, onde o principal objetivo é dividir um conjunto de dados em clusters distintos, sem conhecimento prévio das categorias ou rótulos dos dados. Este algoritmo foi introduzido pela

primeira vez por MacQueen (1967), mas ele pode ser identificado até mesmo em trabalhos anteriores sobre processamento de sinais e otimização matemática (LLOYD, 1982).

$$\sum_{i=1}^{K} \sum_{x \epsilon C_i} ||x - \mu_i||^2 \tag{2.4}$$

O algoritmo visa dividir um conjunto de n observações em K clusters, onde cada observação pertence ao cluster com a média mais próxima, conhecida como centróide. Desta forma, a ideia central por trás do K-means é minimizar a variância dentro do cluster, muitas vezes representada como a soma das distâncias quadradas entre os pontos de dados e seus respectivos centróides de cluster. A função objetivo a ser minimizada é representada pela expressão 2.4, onde C_i representa o i-ésimo cluster, e μ_i é o centróide deste cluster. Assim, o algoritmo K-means envolve os seguintes passos iterativos:

1. Inicialização: são atribuídos K centróides do conjunto de dados;
2. Atribuição: calcular os novos centróides como a média dos pontos atribuídos a cada cluster;
3. Repetir: repete os passos de atribuição e atualização até a convergência, isto é, até que os centróides não mudem significativamente.

A convergência do algoritmo K-means é garantida em um número finito de iterações, pois, a cada passo, o algoritmo reduz a soma das distâncias quadradas. No entanto, a solução final não é garantida como ótima globalmente. O K-means frequentemente converge para um mínimo local, dependendo da inicialização dos centróides Arthur e Vassilvitskii (2007). Para mitigar esse problema, o método de inicialização K-means++ foi proposto, selecionando centróides iniciais de forma mais informada, o que leva a uma convergência mais rápida e a um desempenho aprimorado de agrupamento (ARTHUR; VASSILVITSKII, 2007).

A complexidade computacional do algoritmo K-means é geralmente $O(n \cdot k \cdot d \cdot t)$, onde n é o número de pontos de dados, k é o número de clusters, d é o número de características, e t é o número de iterações. Essa complexidade linear o torna eficiente para grandes conjuntos de dados, embora a natureza iterativa do algoritmo possa torná-lo caro para dados de altíssima dimensionalidade (HARTIGAN; WONG, 1979).

Apesar de sua ampla utilização em diferentes áreas, o algoritmo K-means apresenta algumas limitações. A primeira limitação, é a necessidade de se definir previamente o

número de clusters K, que nem sempre é conhecido. Outra limitação, é a sua característica de assumir que os clusters são esféricos e de tamanho similar, o que pode não ser verdade em muitos conjuntos de dados. Além disso, o K-means é sensível a outliers, pois eles podem afetar significamente o posicionamento dos centróides (NG; HAN, 2002).

2.1.5 Árvores de decisão

O uso formal de árvores de decisão no contexto de ML pode ser creditado a diversos avanços nas áreas de estatística e inteligência artificial durante as décadas de 1960 e 1970. O algoritmo ID3, desenvolvido por J. Ross Quinlan, foi um dos primeiros e mais significativos avanços no campo, introduzido em meados dos anos 1980. Quinlan posteriormente aprimorou o algoritmo, dando origem ao C4.5, uma das versões mais usadas de árvores de decisão Quinlan (1993). A partir desses desenvolvimentos, outros métodos, como o CART (Classification and Regression Trees), foram propostos por Breiman et al. (1984), consolidando o uso de árvores de decisão na estatística e ciência de dados.

O funcionamento de uma árvore de decisão pode ser resumido em três etapas principais: seleção da feature de divisão, divisão dos dados, e avaliação da pureza do nó. Cada uma destas etapas funcionam da seguinte forma:

1. Seleção da feature de divisão: A cada passo da construção da árvore, o algoritmo seleciona a feature que melhor divide o conjunto de dados. A melhor divisão é baseada em métricas como o Ganho de Informação, usado no ID3, ou o Índice de Gini, empregado no CART. O objetivo é escolher a feature que maximiza a separação entre as classes em cada nó da árvore.

2. Divisão dos dados: Uma vez que a feature de divisão é selecionada, o algoritmo divide os dados em subconjuntos de acordo com os valores dessa feature. Essa divisão é recursiva, com cada subconjunto sendo dividido ainda mais até que não seja mais possível realizar uma divisão que melhore o desempenho do modelo.

3. Avaliação da pureza do nó: O algoritmo utiliza métricas como a entropia (para árvores baseadas em ganho de informação) ou a impureza de Gini (para árvores CART) para medir a pureza de um nó. Um nó é considerado puro quando contém amostras pertencentes a uma única classe. A árvore é construída de modo a maximizar a pureza dos nós terminais (folhas), onde são feitas as previsões.

O algoritmo de árvore de decisão é conhecido por sua simplicidade e capacidade de gerar modelos interpretáveis, o que o torna uma escolha natural em diversos domínios.

Porém, um dos problemas comuns nas árvores de decisão é o overfitting — quando o modelo se ajusta demasiadamente aos dados de treinamento, capturando ruídos e anomalias que não generalizam bem para novos dados. Para mitigar isso, é comum aplicar uma técnica chamada poda, que remove ramificações desnecessárias da árvore. A poda pode ser feita durante a construção (poda prévia) ou após a árvore ser totalmente gerada (poda posterior) (BREIMAN et al., 1984).

2.1.6 Random Forest

O Random Forest surgiu a partir do conceito de bagging (bootstrap aggregating), também desenvolvido por Breiman, que consiste em criar vários subconjuntos de treinamento a partir de amostras aleatórias do conjunto original de dados. Assim, cada modelo é treinado de forma independente em um desses subconjuntos, e suas previsões são combinadas para produzir um resultado mais estável Breiman (1996). No caso do Random Forest, essa ideia foi aprimorada ao introduzir um passo adicional de aleatoriedade, pois, além de treinar múltiplas árvores de decisão com diferentes subconjuntos de dados, o algoritmo também seleciona aleatoriamente um subconjunto de features (atributos) para cada árvore. Essa combinação de bagging e aleatoriedade no espaço de features faz com que o Random Forest seja mais robusto e menos suscetível ao overfitting do que uma única árvore de decisão (CUTLER et al., 2007).

O Random Forest é composto por um conjunto de n árvores de decisão, cada uma treinada em um subconjunto aleatório do conjunto de dados original. O algoritmo pode ser descrito em três etapas principais:

1. Amostragem de dados: a partir do conjunto de dados de treinamento original, são gerados múltiplos subconjuntos por meio da técnica de bootstrap. Cada subconjunto contém aproximadamente 63,2% dos dados originais, com reposição, o que significa que alguns dados podem aparecer mais de uma vez e outros podem ser omitidos;

2. Construção das árvores: para cada subconjunto, o Random Forest constrói uma árvore de decisão, mas, diferentemente dos métodos tradicionais, em cada nó da árvore, um subconjunto aleatório de features é selecionado para determinar o critério de divisão Breiman (2001). Isso garante que as árvores sejam diversificadas e não fortemente correlacionadas;

3. Agregação: após a construção de todas as árvores, o Random Forest faz a agregação dos resultados. Para problemas de classificação, a previsão final é baseada no voto

da maioria das árvores (ou seja, a classe mais votada). Para problemas de regressão, a previsão final é a média das saídas das árvores.

Desta forma, a aleatoriedade possui um papel importante na construção das árvores e na seleção das features garantindo que o Random Forest tenha baixa variância e seja altamente resistente ao overfitting, mesmo quando aplicado a grandes volumes de dados e com um alto número de árvores (LIAW; WIENER, 2002).

O algoritmo Random Forest tem sido amplamente utilizado em uma variedade de aplicações de aprendizado supervisionado, tanto em tarefas de classificação quanto de regressão. Suas principais áreas de aplicação incluem: processamento de imagens, detecção de fraude e predição de desempenho de mercado

2.2 Mudanças climáticas

O termo mudança climática refere-se às variações significativas e de longo prazo nas condições meteorológicas de uma região, podendo incluir alterações na temperatura, precipitação e padrões de vento. Essas alterações podem estar ligadas tanto a fatores naturais, como variações na radiação solar, erupções vulcânicas e fenômenos naturais, quanto a ações antrópicas. No entanto, nas últimas décadas, este termo tem sido amplamente utilizado para descrever as mudanças que estão sendo causadas pelas atividades humanas, especialmente pela emissão de gases de efeito estufa (GEE) provenientes da queima de combustíveis fósseis, desmatamento e agricultura intensiva (VANHALA et al., 2023).

Neste contexto, o aquecimento global torna-se um aspecto específico da mudança climática referindo-se ao aumento da temperatura média da superfície terrestre. Este fenômeno é atribuído ao aumento das concentrações de GEE na atmosfera, como o dióxido de carbono, metano e óxidos de nitrogênio. Estes gases, agem na atmosfera do planeta, retendo o calor que deveria ser irradiado de volta para o espaço, causando assim um aumento na temperatura global (VANHALA et al., 2023).

Ainda, de acordo com Calvin et al. (2023), as atividades humanas, principalmente as emissões de GEE, causaram um aumento de 1,1°C na temperatura média global entre 1850 e 2020. Este aumento, embora pareça pequeno, já tem gerado grandes impactos, como o derretimento das calotas polares, aumentando o nível do mar e eventos climáticos extremos, como ondas de calor, secas e tempestades mais intensas. Esses efeitos têm implicações diretas sobre a biodiversidade, a agricultura, os sistemas econômicos e a saúde humana em todo o mundo.

Desta forma, as consequências das mudanças climáticas para a saúde humana, se apresentam de forma multifacetada, afetando diretamente a saúde física e mental da população mundial. Isto pode ser visto em Patz et al. (2005), onde segundo os autores, o aquecimento global contribui para a propagação de doenças infecciosas, alterando os padrões climáticos e afetando a distribuição de novos vetores, como os mosquitos responsáveis pela transmissão de doenças como dengue, malária e zika. Ainda segundo os autores, estudos indicam que, com o aumento das temperaturas, áreas antes não endêmicas se tornam mais suscetíveis à infestação por esses vetores. Estes efeitos também são ressaltados por Hales et al. (2002) que sugere que a mudança nas temperaturas e nas chuvas pode influenciar a sazonalidade e a intensidade de surtos infecciosos.

Os efeitos da mudança climática na saúde humana também são mencionados no trabalho de Willox et al. (2013). Neste caso os autores sugerem que as mudanças climáticas têm um impacto significativo na saúde mental. Devido ao estresse causado por desastres naturais, deslocamento forçado e insegurança econômica, podendo levar a transtornos mentais, como ansiedade e depressão. Além disso, a preocupação contínua com os efeitos das mudanças climáticas pode gerar sentimentos de impotência e desespero nas comunidades afetadas.

O trabalho de Pope 3rd et al. (2009), aborda ainda a poluição do ar causado pelas mudanças climáticas como um outro fator de complicação para a saúde humana. Segundo os autores, o aumento das temperaturas intensificam a formação de poluentes secundários, como o ozônio troposférico, que é prejudicial à saúde respiratória e cardiovascular. Além disso, o aumento da frequência e a intensidade de incêndios florestais em regiões vulneráveis, tem liberando grandes quantidades de material particulado na atmosfera, aumentando os níveis de PM2.5 e agravando condições respiratórias preexistentes, como asma, bronquite e doenças pulmonares obstrutivas crônicas. Esses poluentes podem irritar as vias respiratórias, causando inflamação e diminuindo a função pulmonar. Além disso, os autores destacam que episódios de alta poluição, exacerbados pela mudança climática, têm sido correlacionados com aumento de internações hospitalares por crises respiratórias.

Como pode ser notado nos trabalhos mencionados anteriormente, as mudanças climáticas, intensificadas pela ação humana, têm gerado efeitos profundos e complexos em diversos aspectos da vida na Terra, especialmente na saúde humana. Proporcionando, desta forma, o agravamento de doenças infecciosas e doenças respiratórias, bem como implicações na biodiversidade, agricultura, sistemas econômicos e na saúde pública em todo o mundo.

2.3 Doenças relacionadas ao clima

As doenças relacionadas ao clima são condições de saúde diretamente influenciadas pelas variações climáticas e ambientais. Essas enfermidades podem ser exacerbadas por fenômenos meteorológicos extremos, como ondas de calor, enchentes, secas e furacões, além de mudanças graduais no clima global, como o aquecimento global e a alteração dos padrões de precipitação. Por isso, a relação entre clima e saúde pública é complexa e multifacetada, abrangendo tanto doenças transmissíveis quanto não transmissíveis.

Entre as doenças transmissíveis, destacam-se aquelas que são transmitidas por vetores, como a dengue, malária e zika, cujas ocorrências tendem a aumentar com o aquecimento global. Isso ocorre porque vetores como mosquitos (como por exemplo, o aedes aegypti) encontram ambientes mais propícios para sua proliferação em regiões mais quentes e úmidas (CONFALONIERI et al., 2009). Por outro lado, as doenças respiratórias, como asma e bronquite, podem ser exacerbadas por condições climáticas como altas concentrações de poluentes atmosféricos e aumento de temperatura, que agravam a qualidade do ar (LUBER; PRUDENT, 2009).

As mudanças nos padrões de temperatura e precipitação também afetam a disponibilidade de água potável e o saneamento básico, criando condições ideais para a proliferação de doenças de origem hídrica, como diarreias infecciosas. Estas ondas de calor intensas têm também sido associadas a um aumento na mortalidade por doenças cardiovasculares e respiratórias, particularmente em populações vulneráveis, como idosos e crianças (PATZ et al., 2005).

Além disso, o impacto do clima sobre a segurança alimentar, com prejuízos à produção agrícola e consequente aumento da fome, contribui para o surgimento de condições como a desnutrição e também pode promover deslocamentos populacionais e crises migratórias, levando a um aumento no risco de surtos de doenças em regiões com infraestrutura de saúde precária (WATTS et al., 2018).

Ainda segundo Watts et al. (2018), os eventos climáticos extremos, como furacões, enchentes e secas, também afetam a saúde mental, causando estresse psicológico, transtorno de estresse pós-traumático (TEPT) e ansiedade em indivíduos afetados diretamente por esses eventos.

Como visto anteriormente, as doenças relacionadas ao clima englobam tanto doenças transmissíveis quanto não transmissíveis. Essas enfermidades são influenciadas por fatores ambientais, como variações de temperatura, mudanças nos padrões de precipitação e a

ocorrência de eventos climáticos extremos. A seguir, são descritas com mais detalhes algumas destas doenças, assim como os seus impactos para a saúde humana.

2.3.1 Doenças transmitidas por vetores

As doenças transmissíveis por vetores são infecções que dependem de organismos vetores, como mosquitos, carrapatos e moscas, para a transmissão de patógenos entre humanos ou de animais para humanos. Estas doenças incluem a malária, dengue, chikungunya, Zika, doença de Lyme e febre-amarela, entre outras. A distribuição e a intensidade destas doenças estão fortemente ligadas a fatores ambientais, e o aquecimento global.

O trabalho de Edelson et al. (2022), destaca que a mudança nos padrões de temperatura e precipitação pode expandir os habitats dos mosquitos Anopheles, vetores da malária, resultando em surtos em áreas anteriormente não afetadas. Este fenômeno é exemplificado pela crescente incidência de malária em regiões montanhosas da África Oriental, onde temperaturas mais altas permitem que os mosquitos sobrevivam em altitudes maiores. De forma semelhante, os autores de Vuurst e Escobar (2023) revisaram evidências sobre a relação entre mudanças climáticas e doenças infecciosas, destacando que o aumento das temperaturas facilita a proliferação dos mosquitos Aedes, principais vetores de dengue, chikungunya e Zika. Segundo os autores, a expansão do habitat desses mosquitos tem levado a um aumento significativo na incidência dessas doenças em várias regiões, incluindo áreas urbanas densamente povoadas.

2.3.2 Doenças de origem hídrica

As doenças de origem hídrica representam um desafio para a saúde pública global, especialmente à medida que as mudanças climáticas continuam a impactar a disponibilidade e a qualidade da água. Estas doenças podem ser causadas por diversos patógenos, incluindo vírus, bactérias, protozoários e vermes. Entre as doenças mais comuns estão a cólera, a disenteria bacteriana, a hepatite A e a giardíase. A transmissão destas doenças ocorre principalmente através da ingestão de água contaminada por fezes humanas ou animais, que contêm esses patógenos (NICHOLS et al., 2018).

De acordo com Semenza e Ko (2023), essas doenças contribuem significativamente para a carga de morbidade e mortalidade em regiões com acesso limitado a água potável e saneamento básico adequado. Além disso, as mudanças climáticas e as alterações nos padrões de precipitação podem agravar a situação, aumentando a incidência dessas

doenças.

2.3.3 Doenças respiratórias

As doenças respiratórias são condições que afetam o sistema respiratório, incluindo as vias aéreas, pulmões e estruturas associadas à respiração. Essas doenças podem ser causadas por infecções (bacterianas, virais ou fúngicas), fatores ambientais (poluentes, fumaça, poeira) ou predisposições genéticas, dentre estas doenças, as mais comuns são a asma, bronquite e pneumonia.

Estas doenças podem ser agravadas por condições climáticas como o aumento da poluição do ar e ondas de calor. Segundo Semenza (2020) as ondas de calor podem aumentar a produção de ozônio troposférico, um poluente que irrita as vias respiratórias e agrava condições como asma e bronquite crônica. Ainda segundo os autores, durante eventos de calor extremo, há um aumento nas admissões hospitalares e nas taxas de mortalidade relacionadas a doenças respiratórias.

A poluição do ar também é um dos agravantes para as doenças respiratórias. Esta poluição é resultante de atividades humanas como a queima de combustíveis fósseis e emissões industriais contendo partículas finas (PM2.5) e outros poluentes que podem penetrar profundamente nos pulmões, causando inflamação e danos aos tecidos pulmonares. A exposição prolongada a esses poluentes está associada a um aumento na prevalência de doenças respiratórias crônicas, como a Doença Pulmonar Obstrutiva Crônica (DPOC), e a um maior risco de infecções respiratórias agudas Semenza (2020).

2.3.4 Doenças Cardiovasculares e relacionadas ao calor

As doenças cardiovasculares (DCV) compreendem uma ampla gama de condições que afetam o coração e os vasos sanguíneos. Entre as DCV mais comuns estão a doença arterial coronariana, insuficiência cardíaca, arritmias, hipertensão e doenças das válvulas cardíacas. Essas condições podem levar a eventos graves como ataques cardíacos e acidentes vasculares cerebrais (AVC), representando uma das principais causas de morte no mundo.

Segundo Braunwald (2024), as DCV resultam de uma combinação de fatores genéticos, comportamentais e ambientais. Estes fatores podem resultar no acúmulo de placas de gordura nas paredes das artérias, conhecido como aterosclerose, sendo um processo comum na maioria das doenças cardiovasculares. Essa obstrução reduz o fluxo sanguíneo e pode

levar a isquemia, infarto do miocárdio e outras complicações graves.

Estas doenças também podem ser agravadas por eventos climáticos extremos, como ondas de calor, que possuem um impacto direto na saúde cardiovascular, pois aumento das temperaturas pode provocar desidratação e desequilíbrios eletrolíticos, que podem desencadear arritmias e insuficiência cardíaca. Além disso, a exposição prolongada ao calor aumenta a pressão arterial e a frequência cardíaca, elevando o risco de eventos cardiovasculares Braunwald (2024).

Estudos publicados na JAMA Cardiology e na European Heart Journal mostram que a taxa de mortalidade por DCV aumenta durante períodos de calor extremo. A desidratação e o estresse térmico são fatores críticos que contribuem para o aumento de hospitalizações e mortalidade cardiovascular durante esses eventos climáticos (KAZI et al., 2024).

2.3.5 Saúde mental e impactos psicológicos

As doenças mentais, também conhecidas como transtornos mentais, são condições que afetam o funcionamento normal da mente, emoções e comportamentos de uma pessoa. Essas condições podem variar em gravidade, sintomas e impacto na vida cotidiana, afetando a capacidade de um indivíduo de interagir socialmente e executar tarefas diárias. Entre os tipos mais comuns de doenças mentais estão a depressão, ansiedade, esquizofrenia, transtorno bipolar e transtornos de estresse pós-traumático (TEPT).

Estes transtornos podem ser causados por uma combinação de fatores biológicos, psicológicos e ambientais. Alguns estudos indicam uma correlação entre os desastres ambientais e uma variedade de distúrbios psicológicos, como ansiedade, depressão e TEPT. A literatura atual evidencia que o aumento da frequência e intensidade desses eventos está diretamente ligado a efeitos negativos na saúde mental, agravando condições pré-existentes e gerando novos transtornos emocionais (BERRY et al., 2009; CLAYTON et al., 2017).

Segundo Berry et al. (2009), indivíduos expostos a eventos climáticos extremos, como inundações, incêndios florestais e secas, enfrentam maiores riscos de desenvolver distúrbios emocionais. Estes distúrbios, podem durar meses ou até anos, e são mais prevalentes em populações de baixa renda, com menor acesso a serviços de saúde e suporte social. Além disso, certos grupos, como crianças, mulheres e populações indígenas, são particularmente vulneráveis devido a fatores socioeconômicos e biológicos que agravam os impactos psicológicos (CLAYTON et al., 2017).

Portanto, os efeitos psicológicos das mudanças climáticas, não podem ser dissociados do contexto social e ambiental mais amplo, tornando necessária uma abordagem holística para mitigar esses impactos. Intervenções que integram apoio psicológico com ações ambientais são importantes para lidar com as crises atuais e futuras causadas pelas mudanças climáticas (CLAYTON et al., 2017).

2.4 Sistemas de saúde e sua atuação no tratamento de doenças

Os sistemas de saúde são formados por estruturas organizacionais e operacionais que buscam promover, restaurar e manter a saúde das populações. Eles envolvem uma ampla gama de atores, incluindo governos, prestadores de serviços de saúde, indústrias farmacêuticas, organizações internacionais e comunidades. Desta forma, o funcionamento de cada sistema de saúde depende de fatores econômicos, culturais e políticos de cada país, o que resulta em variações na forma como esses sistemas são estruturados e administrados Frenk (2010).

No entanto, ainda é possível notar semelhanças entre os sistemas de saúde de diferentes países. Estas semelhanças podem ser percebidas em Tikkanen et al. (2020), que faz uma comparação entre os seguintes aspectos apresentados a seguir:

- **Cobertura universal:** a cobertura universal de saúde é uma das principais características que une sistemas de saúde de muitos países desenvolvidos. Em nações como Austrália, Brasil, Canadá, Reino Unido, e grande parte da Europa, todos os residentes têm acesso a cuidados médicos essenciais, independentemente de sua condição financeira. Isso significa que todos têm o direito de acessar tratamentos médicos de qualidade, incluindo consultas, exames e tratamentos hospitalares, sem a preocupação de que esses serviços resultem em dívidas ou exclusão por incapacidade de pagamento. Além disso, a implementação da cobertura universal é vista como uma forma de garantir equidade, ao assegurar que todos os cidadãos, independentemente de sua classe social ou local de residência, tenham acesso ao mesmo padrão de atendimento. A universalidade visa evitar que os mais vulneráveis fiquem sem atendimento, promovendo uma distribuição mais justa dos serviços de saúde. A cobertura universal também proporciona um sistema de proteção social que previne catástrofes financeiras derivadas de emergências de saúde.

- **Financiamento via tributos:** nos países com sistemas universais, o financiamento

das despesas de saúde é, em sua maior parte, oriundo de tributos ou receitas fiscais gerais. Alguns exemplos incluem Austrália, Brasil, e Inglaterra, onde o governo desempenha um papel central na arrecadação e alocação de recursos para o setor de saúde. Esse modelo visa garantir que os serviços de saúde sejam financiados de forma equitativa, redistribuindo recursos coletados através de impostos, especialmente os progressivos, em que cidadãos com maiores rendimentos contribuem mais. Desta forma, o financiamento público torna o acesso aos serviços de saúde mais acessível, ao eliminar ou reduzir os custos diretos para os pacientes no ponto de atendimento. Isso cria um cenário em que o atendimento médico básico, incluindo visitas ao médico e hospitalizações, seja acessível, independentemente da renda do indivíduo. Esse modelo também permite que o governo mantenha controle sobre os gastos de saúde e regule melhor os preços de medicamentos e serviços, como ocorre no Reino Unido e Canadá.

- **Participação do governo na regulação:** em vários países, o governo desempenha um papel decisivo na regulação tanto do sistema público quanto do privado de saúde. Isso inclui a definição de normas de qualidade, a regulação dos preços dos serviços médicos e medicamentos, bem como a criação de incentivos para o uso de serviços preventivos. A regulação governamental busca garantir que o sistema de saúde funcione de forma eficiente e segura, assegurando que os serviços de saúde sejam prestados com qualidade e que os custos sejam controlados. Essa regulação é essencial para manter um equilíbrio entre os setores público e privado, além de evitar a exploração financeira no setor da saúde. Por este motivo, em muitos países, como na França, Austrália e Brasil, o governo define limites para os preços cobrados por serviços de saúde e medicamentos, tornando o atendimento acessível para a população em geral. Além disso, políticas de incentivo a cuidados preventivos e programas de saúde pública ajudam a reduzir a carga sobre o sistema de saúde ao prevenir doenças antes que elas se tornem graves.

- **Segmentação em cuidados primários e especializados:** os sistemas de saúde globais, em geral, adotam uma estrutura que prioriza os cuidados primários como porta de entrada para o sistema de saúde. Isso significa que os pacientes geralmente devem consultar médicos generalistas, que são os responsáveis por fornecer os primeiros diagnósticos e encaminhar os pacientes para especialistas, quando necessário. Essa segmentação entre cuidados primários e especializados é comum em sistemas como o da Austrália e Brasil, onde o atendimento primário é acessível, enquanto o atendimento especializado é obtido com base em referências médicas. Assim, a

segmentação em cuidados primários e especializados visa otimizar os recursos do sistema de saúde, garantindo que apenas os casos mais complexos sejam encaminhados a especialistas, enquanto as condições mais comuns são tratadas de maneira eficiente pelos médicos de atenção primária.

- **Seguro privado e complementar:** em muitos países com sistemas universais de saúde, ainda há espaço para seguros privados complementares, que cobrem serviços adicionais não contemplados pelo sistema público. Isso é observado em países como França, Austrália e Brasil, onde os seguros privados atuam como uma maneira de aliviar gargalos no sistema público, permitindo aos pacientes maior flexibilidade e rapidez no acesso a determinados serviços de saúde, como cirurgias eletivas, tratamentos odontológicos e consultas com especialistas. O seguro privado complementar não substitui o sistema público, mas oferece uma camada adicional de cobertura para aqueles que podem pagar por serviços extras. Em países como a França, por exemplo, os seguros complementares são usados para cobrir coparticipações e serviços que não são integralmente financiados pelo sistema público. No Brasil, uma parcela da população de classe média e alta opta por planos privados para obter um atendimento mais rápido, enquanto ainda mantém o direito de acesso ao Sistema Único de Saúde (SUS).

Esta estrutura apresenta parte da complexidade dos sistemas de saúde comum entre vários países e o quão desafiador ela pode se tornar dentro de um contexto de crises econômicas e de doenças emergentes. Estes desafios são mencionados por Whiteford e Vindrola-Padros (2015), destacando que, apesar de haver uma tendência mundial de aumentar os gastos com serviços de saúde, os governos têm encontrado dificuldades em fornecer assistência médica de boa qualidade a uma população em mudança e expansão, dentro de um contexto de recursos escassos e crises econômicas. Ainda segundo Whiteford e Vindrola-Padros (2015), o aumento dos fluxos internacionais de pessoas, o agravamento das condições de vida, as mudanças climáticas severas e o uso excessivo de antibióticos contribuíram para o surgimento de novas doenças infecciosas como SARS, gripe aviária e H1N1 bem como o ressurgimento de doenças anteriormente erradicadas, representando novos desafios para sistemas de saúde já sobrecarregados.

Além disso, apesar dos esforços, existem barreiras para a adaptação eficaz dos sistemas de saúde às mudanças climáticas. Entre os principais obstáculos identificados estão a falta de políticas climáticas adequadas, deficiências nos sistemas de preparação para desastres, escassez de recursos, implementação e avaliação deficientes de políticas, além

da percepção de baixo risco por parte da população e da falta de expertise e dados científicos. Essas barreiras comprometem a capacidade de resposta eficaz dos sistemas de saúde, especialmente em países em desenvolvimento, onde os sistemas de saúde são mais frágeis. (ANSAH et al., 2024).

De acordo com Ansah et al. (2024), para superar essas dificuldades e enfrentar os impactos cada vez mais frequentes e severos das mudanças climáticas, é fundamental fortalecer os sistemas de saúde de forma a resistir às crises climáticas, além de garantir a resiliência dos profissionais de saúde e das populações vulneráveis. Esses esforços são alinhados ao quadro de resiliência dos sistemas de saúde da Organização Mundial da Saúde, que enfatiza a necessidade de políticas integradas, governança eficiente e o fortalecimento da capacidade institucional para lidar com as incertezas climáticas.

2.5 Revisão de Mapeamento

A revisão de Mapeamento (RM), é uma metodologia que se distingue de outras revisões pela sua abrangência ao tratar de tópicos amplos dentro de um campo de pesquisa, como por exemplo, a engenharia de software. Esta metodologia é utilizada como uma ferramenta exploratória que busca categorizar e classificar os estudos publicados, proporcionando uma visão geral sobre quais áreas são mais estudadas, quais métodos são mais utilizados e onde existem lacunas de pesquisa . Sendo assim, o principal objetivo de uma RM é fornecer uma visão panorâmica de uma área de conhecimento, que pode servir como base para investigações mais focadas, como revisões sistemáticas detalhadas ou novas pesquisas empíricas. Desta forma, uma RM não possui como objetivo responder a perguntas de pesquisa específicas, mas sim em organizar a literatura existente, permitindo que os pesquisadores identifiquem temas recorrentes, tendências e lacunas que precisam ser exploradas (KITCHENHAM et al., 2011).

As diferenças entre uma RSL e uma RM são abordadas em Kitchenham et al. (2011) e apresentadas na Tabela 1 onde é feita uma comparação entre os dois métodos de pesquisa em cada uma das etapas do processo.

Desta forma, como pode ser observado na Tabela 1, a metodologia de uma RM segue etapas semelhantes às de uma revisão sistemática da literatura (RSL), mas com diferenças importantes. Segundo Kitchenham et al. (2011), essas etapas incluem:

- **Definição das questões de pesquisa:** Normalmente, essas questões são mais amplas e podem incluir perguntas como "Quais subtemas estão sendo investigados?"ou

Tabela 1: Comparação entre tipos de revisão.

Elementos da RSL	Revisão de Mapeamento	Revisão Sistemática
Objetivos	Classificação e análise de temas da literatura dentro do tópico de engenharia de software	Identificar as melhores práticas no que diz respeito a procedimentos, tecnologias, métodos ou ferramentas específicas, agregando informações de estudos comparativos
Questões de Pesquisa	Genérico – relacionado às tendências de pesquisa. Da forma: quais pesquisadores, quanta atividade, que tipo de estudos, etc.	Específico – relacionado aos resultados de estudos empíricos. Exemplo: A tecnologia/método A é melhor ou não que B?
Processo de Busca	Definido pela área do tópico	Definidos pelas questões de pesquisa, as quais identificam as tecnologias específicas a serem investigadas
Escopo	Amplo – todos os artigos relacionados a uma área temática são incluídos, mas apenas dados de classificação sobre eles são coletados	Focado – apenas artigos empíricos relacionados a uma questão de pesquisa específica são incluídos e informações detalhadas sobre resultados de pesquisa individuais são extraídas de cada artigo
Estratégia de Busca	Muitas vezes menos rigoroso se apenas as tendências de investigação forem de interesse, por exemplo os autores podem pesquisar apenas um conjunto específico de publicações, restringir limitar-se a artigos de periódicos ou restringir-se a um ou dois bibliotecas digitais	Extremamente rigoroso – todos os estudos relevantes devem ser encontrados. Geralmente na RSL os autores precisam utilizar outras técnicas além da simples pesquisa de fontes de dados, como olhar para as referências em estudos primários identificados e/ou abordar pesquisadores da área para descobrir se eles estão realizando novas pesquisas na área
Avaliação de Qualidade	Não é essencial. Se torna complicado pela natureza inclusiva da pesquisa que pode incluir estudos teóricos, bem como estudos empíricos, dificultando a avaliação da qualidade dos estudos primários	Importante para garantir que os resultados sejam baseados em evidências da melhor qualidade
Resultados	Um conjunto de artigos relacionados a uma área temática categorizada em uma variedade de dimensões e contagens do número de artigos em diversas categorias	Os resultados dos estudos primários são agregados para responder às questão(ões) de pesquisa específica(s), possivelmente com qualificadores (por exemplo, os resultados se aplicam apenas para novatos)

Fonte: Adaptado de Kitchenham et al. (2011).

"Quais métodos de pesquisa são utilizados com mais frequência?";

- **Busca de estudos primários:** assim como em uma RSL, uma busca extensa é realizada em bases de dados acadêmicas para identificar todos os estudos relevantes dentro de um campo;
- **Seleção de estudos relevantes:** nesta etapa, critérios de inclusão e exclusão são aplicados para garantir que apenas estudos que respondam às perguntas de pesquisa amplas sejam considerados;

- **Classificação dos estudos:** nesta etapa, os estudos são classificados com base em vários parâmetros, como ano de publicação, tipo de estudo (teórico, empírico, etc.), métodos utilizados, e tópicos abordados;
- **Visualização dos resultados:** esta etapa é importante em uma RM pois permite a visualização dos resultados em gráficos e tabelas que mostrem a distribuição da literatura em diferentes categorias;

Como se pode notar, a maior diferença entre uma RM e uma RSL está relacionada ao escopo e ao resultado esperado. Enquanto uma RSL busca responder a perguntas específicas com base na agregação de resultados de estudos empíricos, uma RM possui uma abordagem mais geral e descritiva visando organizar a literatura de uma área ampla, mas geralmente não avalia a qualidade ou os resultados empíricos dos estudos de forma tão detalhada.

Um exemplo dessa distinção está na forma como as questões de pesquisa são formuladas. Em uma RSL, a pergunta pode ser algo como: "O algoritmo A é mais eficaz que o algoritmo B?". Já em uma RM, a pergunta seria mais ampla, do tipo: "Quais algoritmos estão sendo mais utilizados na correlação entre doenças e mudanças climáticas?". Dessa forma, uma RM pode ser vista como uma etapa inicial que oferece uma visão geral, enquanto a RSL aprofunda-se em aspectos específicos.

De acordo com Kitchenham et al. (2011), as RMs são importantes porque economizam tempo e esforço ao fornecer uma base sólida para a pesquisa subsequente, permitindo que os pesquisadores identifiquem áreas sub exploradas e façam uma análise preliminar de como um campo está evoluindo e possibilitando uma priorização mais estratégica de novas pesquisas. Além disso, são ferramentas úteis para novos pesquisadores ou estudantes de doutorado, pois proporcionam uma visão estruturada da literatura e ajudam a entender os principais desafios e métodos na área.

Por outro lado, os limites de uma RM incluem o fato de que nela não se aprofunda nos resultados empíricos dos estudos analisados, o que pode limitar sua utilidade quando se busca recomendações práticas imediatas. Outra limitação é que a qualidade da RM depende diretamente da completude e rigor da busca e classificação dos estudos (KITCHENHAM et al., 2011).

O estudo de Sjoeberg et al. (2005) é um exemplo de RM bem-sucedida, que identificou e categorizou grandes volumes de pesquisa em estimativa de custos de software e experimentos centrados no ser humano, respectivamente. Esses estudos forneceram um

ponto de partida valioso para revisões sistemáticas e novas pesquisas empíricas nas áreas mapeadas.

Já o trabalho Apetorgbor et al. (2024) oferece uma análise abrangente das técnicas de Machine Learning (ML) aplicadas para a detecção e predição da malária. Neste trabalho, a utilização da RM ajudou a identificar os avanços mais recentes na aplicação de técnicas de ML aplicadas na detecção da doença, incluindo classificadores supervisionados, como Máquinas de Vetores de Suporte (SVM), k-NN (k-nearest neighbors) e Redes Neurais Convolucionais (CNN), que são utilizados para a análise de imagens de esfregaço sanguíneo e para a detecção automatizada do parasita da malária.

Diferente de uma RSL, que busca responder a perguntas específicas e avaliar a qualidade da evidência de forma rigorosa, a RM é utilizada em cenários onde a evidência é emergente e ainda não há questões específicas bem definidas. Neste contexto, em Munn et al. (2018) os autores fornecem orientações sobre quando utilizar uma RSL ou uma RM de acordo com o objetivo do trabalho. Segundo os autores, uma RM é apropriada quando a pesquisa possui os seguintes objetivos:

- Identificar os tipos de evidência disponíveis em um campo;
- Classificar conceitos ou definições importantes;
- Examinar como a pesquisa é conduzida em um determinado tema;
- Identificar características ou fatores chave relacionados a um conceito;
- Servir como precursor para uma revisão sistemática;
- Identificar e analisar lacuna no conhecimento;

Por outro lado, segundo Munn et al. (2018), uma RSL é mais adequada para:

- Confirmar ou refutar práticas autuais com base em evidências;
- Identificar variações em práticas e resolver conflitos de evidência;
- Gerar declarações para orientar a tomada de decisões clínicas e políticas;

O trabalho de Munn et al. (2018) menciona ainda que uma RM não pode ser confundida com uma revisão tradicional, pois revisões tradicionais não seguem um protocolo padronizado e estruturado, sendo mais flexíveis e baseadas na experiência e conhecimento

prévio do autor. Assim, este tipo de revisão pode não incluir critérios de inclusão/exclusão bem definidos, e o processo de busca e seleção dos estudos pode ser menos rigoroso. Portanto, a RM proporciona uma alternativa mais estruturada e abrangente em comparação a uma revisão tradicional, oferecendo maior rigor metodológico, transparência e reprodutibilidade.

Como visto anteriormente, a RM é uma ferramenta útil para se obter uma visão panorâmica sobre áreas de conhecimento amplas e emergentes, como a engenharia de software e a aplicação de técnicas de Machine Learning. Ao categorizar e classificar estudos, a RM oferece uma base para futuras investigações mais focadas e detalhadas, como as RSLs. Embora ambas sigam etapas semelhantes, suas finalidades são distintas: a RM é exploratória, identificando tendências, lacunas e metodologias, enquanto a RSL visa responder perguntas de pesquisa específicas e fornecer recomendações práticas com base em evidências empíricas. Contudo, o sucesso de uma RM depende da abrangência e rigor da busca de estudos, destacando sua importância como etapa inicial de pesquisa.

2.5.1 Parsifal

O protocolo de uma RM, estabelece um processo rigoroso e estruturado com a finalidade de garantir a transparência, a reprodutibilidade e a objetividade na revisão da literatura, de forma a reduzir vieses e garantir que os resultados reflitam com precisão o estado atual do conhecimento sobre o tema de pesquisa (KITCHENHAM, 2004). Dada a sua complexidade, torna-se importante o uso de ferramentas especializadas para apoiar esse processo de forma a facilitar a organização e a gestão de grandes volumes de dados, além de garantir maior precisão e consistência ao longo das etapas, minimizando erros manuais e vieses de interpretação.

Neste contexto, existem diferentes ferramentas que visam apoiar os autores na gestão e condução de uma RM, automatizando tarefas repetitivas como seleção de estudos, remoção e duplicatas. O trabalho de Stefanovic et al. (2021), faz uma comparação detalhada sobre três destas ferramentas, sendo elas: Sesra, Relis e Parsifal. Nesta avaliação, os autores utilizam a metodologia DESMET, desenvolvida por Marshall et al. (2014), que é utilizada para avaliar métodos e ferramentas de engenharia de software, considerando aspectos quantitativos e qualitativos. Neste caso, os autores utilizaram o método de análise por features, por ser mais adequado para avaliar as ferramentas no contexto de suporte à RM. De acordo com a avaliação feita pelos autores, a ferramenta ReLis obteve o maior desempenho com uma pontuação de 77%, enquanto a ferramenta SESRA obteve

pontuação de 72% e a ferramenta Parsifal ficou com uma pontuação de 68%.

Apesar de sua baixa pontuação em relação as outras duas ferramentas comparadas por Marshall et al. (2014), o Parfifal foi escolhido como ferramenta de apoio para esta RM, devido à familiaridade com sua utilização.

O Parsifal é uma ferramenta online, que auxilia a organização e documentação dos processos de uma RM, oferecendo funcionalidades para o registro das perguntas de pesquisa, criação das estratégias de busca, escolha das fontes de dados e critérios de inclusão e exclusão. Além disso, a ferramenta fornece funcionalidades que permite avaliar a qualidade dos artigos encontrados, bem como a criação de formulário para a coleta de dados. Já durante a fase de execução, o Parsifal facilita o processo ao permitir a importação de arquivos BibTeX, um formato utilizado para organizar e gerenciar referências bibliográficas, além de auxiliar na identificação de estudos duplicados que podem ter sido encontrados em diferentes bases de dados, garantindo que cada estudo seja considerado apenas uma vez (Parsifal, 2024). Estas funcionalidades são descritas a seguir:

2.5.1.1 Planejamento

Em sua etapa de planejamento, o Parsifal permite o preenchimento das informações de protocolo, definição da avaliação de qualidade e definição do formulário para extração de dados. A descrição da definição das etapas de planejamento no Parsifal são apresentadas a seguir:

- **Definição dos objetivos:** nesta seção são definidos os objetivos da RM. Na ferramenta Parsifal esta sessão está identificada como Objectives;
- **Definição PIPOC:** é neste formulário onde são definidas a população, intervenção, comparação, outcomes e contexto da RM. Estas informações são baseadas no objetivo da RM e auxiliam na definição da string de busca;
- **Elaboração da questão de pesquisa:** neste formulário são definidas as questões de pesquisa;
- **Definição de palavras-chave:** nesta sessão, são definidas as palavras-chave e os seus sinônimos, onde cada uma destas palavras pode estar associada a um termo PICOC. Também é possível importar estas palavras de forma automática dos termos PICOC anteriormente definidos. Estas palavras-chave são utilizadas na definição automática da string de busca;

- **Definição das fontes:** nesta sessão, é possível definir as bases de dados de onde serão extraídos os trabalhos primários. Por meio do formulário é possível informar o nome e a url da base de dados;
- **Definição dos critérios de seleção:** os critérios de inclusão e exclusão de trabalhos primários são definidos nesta sessão;
- **Definição dos dados a serem extraídos:** nesta sessão o Parsifal, permite definir quais dados se deseja extrair dos trabalhos selecionados. Estas informações aqui definidas são utilizadas na etapa de extração de dados;

2.5.1.2 Condução

A etapa de condução na ferramenta Parsifal, refere-se a condução da RM, permitindo realizar a busca, seleção, extração de dados e análise de estudos. Essa etapa envolve várias sub-etapas que visam garantir a coleta e análise rigorosa de informações relevantes para a pesquisa. A seguir estão as principais atividades desta fase:

- **Busca de estudos:** nesta etapa, é realizada a busca nos bancos de dados previamente definidos durante a etapa de planejamento, utilizando as estratégias de busca formuladas. O Parsifal permite nesta etapa a definição de strings de busca específicas para cada fonte de busca;
- **Seleção de estudos:** nesta etapa, os estudos identificados anteriormente são filtrados com base nos critérios de inclusão e exclusão definidos na etapa de planejamento;
- **Extração de dados:** aqui os dados dos estudos selecionados são extraídos para posterior análise. Esta tarefa é realizada por meio dos formulários de extração de dados personalizados para organizar estas informações de maneira eficiente.
- **Síntese dos resultados:** A partir dos dados extraídos, o revisor pode realizar uma síntese qualitativa ou quantitativa dos resultados dos estudos. O Parsifal apoia o rastreamento e organização dessas informações.

2.5.1.3 Documentação

Ao final da revisão sistemática, o Parsifal auxilia na geração de um relatório final incluindo os resultados da análise, a síntese dos estudos e as conclusões. Esse relatório

é o produto final da RM e contem todas as informações necessárias para que outros pesquisadores possam entender o processo e replicá-lo, se necessário.

Como se pode notar, o Parsifal é uma ferramenta que otimiza e estrutura cada etapa da RM, promovendo maior controle, organização e transparência no processo de síntese do conhecimento.

3 PROCESSO DE REVISÃO DE MAPEAMENTO

Neste capítulo, são apresentadas cada etapa do processo da revisão de mapeamento empregada neste trabalho, baseando-se no método apresentado por Kitchenham et al. (2011).

3.1 Planejamento da Revisão de Mapeamento (RM)

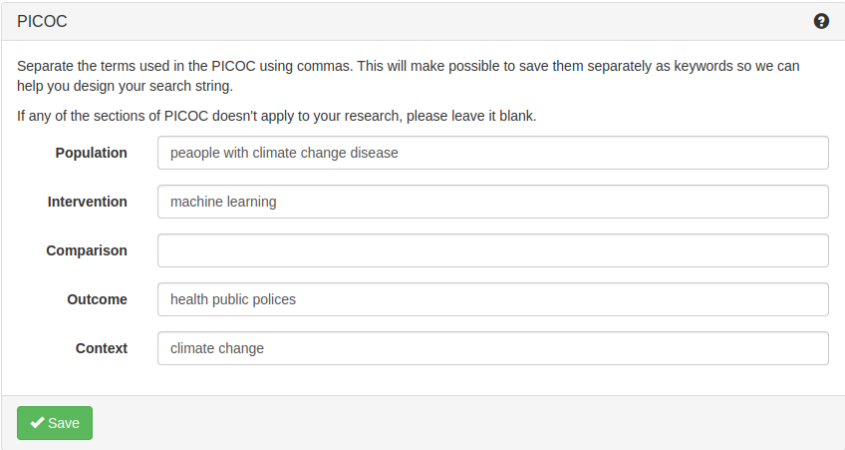
O planejamento uma de RM consiste na definição de um protocolo bem definido, padronizado e estruturado, de forma que a sua aplicação assegure a transparência e reprodutibilidade do processo (MUNN et al., 2018). Por este motivo, a etapa de planejamento torna-se importante pois, serve como base para a condução de todo o processo da RM

Sendo assim, a primeira etapa do planejamento consiste na elaboração de strings de busca adequadas que permitam encontrar artigos que sejam relevantes ao tópico de pesquisa deste trabalho. Desta forma, com o objetivo de garantir as strings de busca mais apropriadas, foi utilizado o protocolo PICOC (Population, Intervention, Comparison, Outcomes, Context) proposto em Kitchenham e Charters (2007), porém, por se tratar de uma RM, o item Comparação do protocolo PICOC não será utilizado. Este protocolo auxilia na extração de palavras-chave baseando-se no tópico de pesquisa. A sua utilização com o auxílio do Parsifal é apresentada na Figura 6, onde, baseando-se no tópico de pesquisa, foram extraídas as palavras-chave relacionadas com cada item do protocolo. Estas palavras-chave são posteriormente organizadas em uma lista com seus respectivos sinônimos conforme a Tabela 2 e também a Figura 7, onde é criada uma lista de palavras-chave e seus sinônimos no Parsifal.

Isto possibilita documentar todo o processo de mapeamento bem como automatizar algumas etapas do processo, como por exemplo, a geração da string de busca, que é sugerida pelo Parsifal baseando-se nas palavras-chave e seus sinônimos.

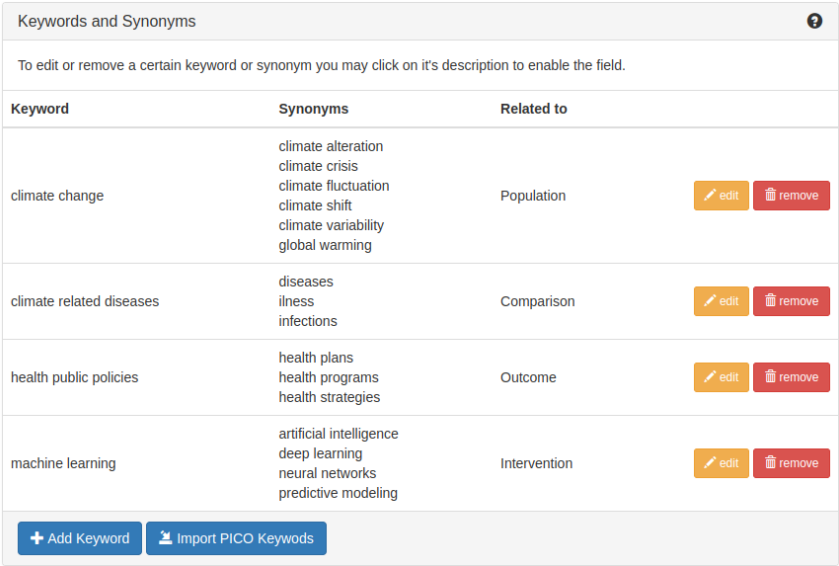
Após extraída a lista das palavras-chave e seus sinônimos, foi realizada no Google

Figura 6: Separação do tema em palavras-chave com método PICOC



Fonte: Autor.

Figura 7: Gerenciamento das palavras-chave e sinônimos no Parsifal



Fonte: Autor.

Tabela 2: Palavras-chave e sinônimos

climate change	**machine learning**	**climate related diseases**	**public health polices**
climate alteration	artificial intelligence	climate impacted diseases	health plans
climate crisis	deep learning	climated related health outcomes	health programs
climate fluctuation	neural networks	climate related illness	health strategies
climate shift	predictive modeling	climate related infections	
climate variability		climate sensitive diseases	
global warming			

Fonte: Autor.

Tabela 3: Avaliação de aderência

Tema	Palavra-chave	Número de retornos	Representatividade (%)
Climate change	climate change	3.600.000	43,99
	climate alteration	2.510	0,03
	climate crisis	364.000	4,45
	climate fluctuation	53.200	0,65
	climate shift	14.800	0,18
	climate variability	1.190.000	14,54
	global warming	2.960.000	36,17
Machine learning	machine learning	5.820.000	23,52
	artificial Intelligence	4.760.000	19,23
	deep learning	6.050.000	24,44
	neural networks	4.810.000	19,43
	predictive modeling	3.320.000	13,37
Climate Related Diseases	climate related diseases	5710000	47,50
	climate-related health outcomes	4.840.000	40,27
	climate-related illness	1.850.000	15,39
	climate-related infections	2.470.000	20,55
	climate-sensitive diseases	2.860.000	23,79
Public health polices	public health polices	4.250.000	15,41
	health plans	7.940.000	28,79
	health programs	8.670.000	31,44
	health strategies	6.720.000	24,37

Fonte: Autor.

Scholar uma busca preliminar para avaliação da aderência dos temas selecionados. As taxas de aderência são apresentadas na Tabela 3, onde é possível notar que não foram detectadas palavras-chave com representatividade nula, o que resultaria em sua eliminação.

Também é possível notar quais foram as palavras-chave com a maior representatividade dentro de cada tema, sendo: I) climate change; II) deep learning; III) climate related diseases e; IV) health programs. Porém, por se tratar uma RM, este trabalho considerou a utilização de todas as palavras-chave, objetivando um estudo mais abrangente se comparado a uma RSL (KITCHENHAM et al., 2011).

A eficácia de uma string de busca depende da escolha cuidadosa das palavras-chave e da compreensão das necessidades da pesquisa. Além disso, são utilizados operadores booleanos com o objetivo de efetuar a combinação entre os termos.

Desta forma, com o objetivo de obter uma string de busca abrangente, algumas das palavras-chave anteriores foram simplificadas. Mais especificamente as palavras-chave relacionadas ao tema *Climate Related Diseases*, reduzindo para três o número de palavras-chave da seguinte forma: diseases; illness; e infections. A decisão para esta mudança foi tomada após um teste utilizando a string de busca originalmente gerada, o que resultou em um retorno de apenas cinco trabalhos para a base de dados IEEE Digital Library.

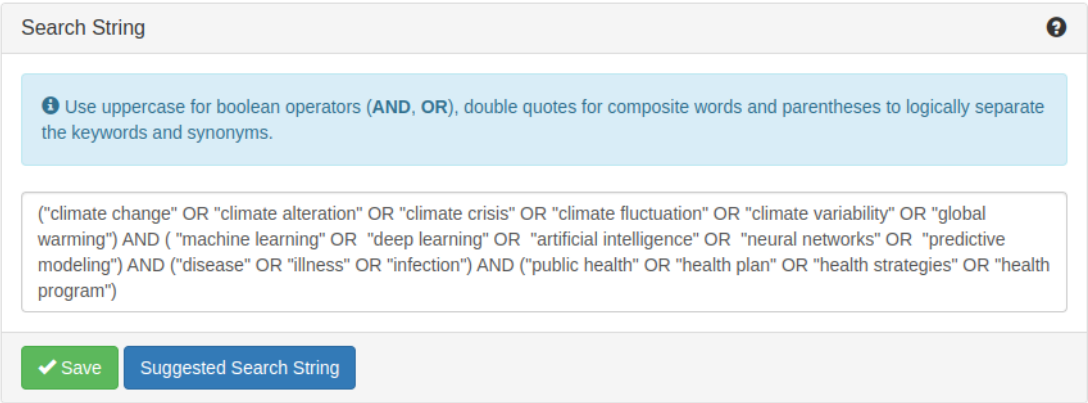
A partir de então, com a definição das palavras-chave e seus sinônimos, foi possível gerar de forma automática com o auxílio do Parsifal a seguinte string de busca:

- **SB1:** (climate change OR climate alteration OR climate crisis OR climate fluctuation OR climate shift OR climate variability OR global warming) AND (machine learning OR artificial intelligence OR deep learning OR neural networks OR predictive modeling) AND (diseases OR infections OR illness) AND (health public polices OR health plans OR health programs OR health strategies)

A criação automática da string de busca por meio do Parsilfal é apresentada na Figura 8, onde também é possível notar a utilização dos operadores booleanos AND e OR, que foram utilizados para realizar a combinação entre os termos, garantindo assim, resultados relevantes e ao mesmo tempo abrangentes.

Outro passo importante na realização da busca por trabalhos relevantes é a definição das bases de dados que serão utilizadas. Baseando-se no tópico de pesquisa deste trabalho, foram escolhidas as bases de dados: IEEE Digital Library, ACM, PubMed, Scopus e Proquest. Na escolha destas bases de dados também foram consideradas a possibilidade de acesso aos artigos por meio da Universidade de São Paulo, além da possibilidade de se utilizar strings de busca para filtrar os resultados. No entanto, nem todas as bases compartilham a mesma sintaxe para strings de busca, por este motivo tornou-se necessário adaptar a string de busca SB1 para utilização em cada uma das bases escolhidas como

Figura 8: Criação da string de busca por meio do Parsifal



Fonte: Autor.

mostrado na Tabela 4.

Tabela 4: Adaptação da string de busca para cada base de dados.

Base	**String de Busca**
IEEE Digital Library	("All Metadata":"climate change"OR "All Metadata":"climate alteration"OR "All Metadata":"climate crisis"OR "All Metadata": "climate fluctuation"OR "All Metadata":"climate variability"OR "All Metadata":"global warming") AND ("All Metadata": "machine learning"OR "All Metadata": "deep learning"OR "All Metadata": "artificial intelligence"OR "All Metadata": "neural networks"OR "All Metadata": "predictive modeling") AND ("All Metadata":"disease"OR "All Metadata": "illness"OR "All Metadata":"infection") AND ("Full Text & Metadata":"public health"OR "Full Text & Metadata":"health plan"OR "Full Text & Metadata":"health strategies"OR "health program")
ACM	AllField:("climate change"OR "climate alteration"OR "climate crisis"OR "climate fluctuation"OR "climate shift"OR "climate variability"OR "global warming") AND AllField:("machine learning"OR "artificial intelligence"OR "deep learning"OR "neural networks"OR "predictive modeling") AND AllField:("public health"OR "health plan"OR "health program"OR "health strategies") AND AllField:("disease"OR "illness"OR "infection")
PubMed	("climate change"[Title] OR "climate alteration"[Title] OR "climate crisis"[Title] OR "climate fluctuation"[Title] OR "climate shift"[Title] OR "climate variability"[Title] OR "global warming"[Title]) AND ("machine learning"[Title] OR "artificial intelligence"[Title] OR "deep learning"[Title] OR "neural networks"[Title] OR "predictive modeling"[Title]) AND (disease OR infection OR illness) AND ("public health"OR "health plans"OR "health program"OR "health strategies")
Scopus	TITLE("climate change"OR "climate alteration"OR "climate crisis"OR "climate fluctuation"OR "climate variability"OR "global warming") AND TITLE("machine learning"OR "deep learning"OR "artificial intelligence"OR "neural networks"OR "predictive modeling") AND ALL("disease"OR "illness"OR "infection") AND ALL("public health"OR "health plan"OR "health strategies"OR "health program")
ProQuest	TITLE("climate change"OR "climate alteration"OR "climate crisis"OR "climate fluctuation"OR "climate variability"OR "global warming") AND TITLE("machine learning"OR "deep learning"OR "artificial intelligence"OR "neural networks"OR "predictive modeling") AND ("disease"OR "illness"OR "infection") AND ("public health"OR "health plan"OR "health strategies"OR "health program")

Fonte: Autor.

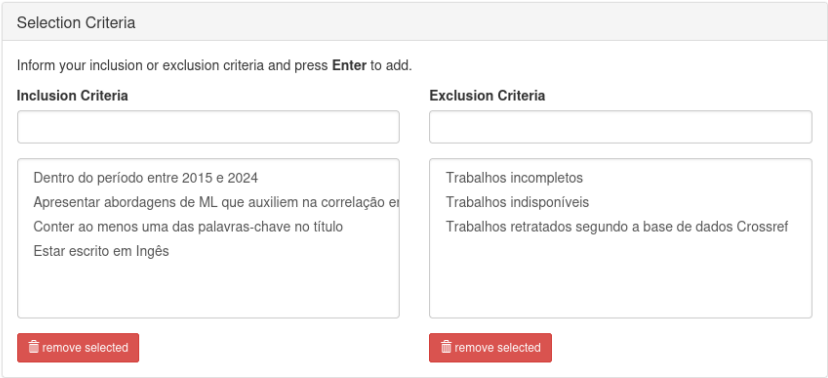
Com a definição das strings de busca e das bases de dados a serem utilizadas, a próxima etapa do planejamento consiste na definição dos critérios de inclusão e exclusão

Tabela 5: Critérios de Inclusão e Exclusão

Critério	**Descrição**
Inclusão	Estar escritos em inglês
	Dentro do período entre 2015 e 2024
	Conter ao menos uma das palavras-chave no título
	Apresentar abordagens de ML que auxiliem na correlação entre mudanças climáticas e doenças relacionadas ao clima
Exclusão	Trabalhos incompletos
	Trabalhos indisponíveis
	Trabalhos retratados segundo a base de dados Crossref

Fonte: Autor.

Figura 9: Critérios de Inclusão e Exclusão na ferramenta Parsifal



Fonte: Autor.

dos resultados. Estes critérios, possuem como objetivo identificar os estudos primários que fornecem evidências diretas sobre as questões de pesquisa, e devem ser definidos durante a elaboração do protocolo, embora possam ser refinados durante o processo de busca, na etapa de condução (KITCHENHAM; CHARTERS, 2007).

Com o objetivo de identificar os estudos primários mais relevantes para os objetivos específicos deste trabalho, foram definidos os critérios de inclusão e exclusão, apresentados na Tabela 5. Estes critérios foram também adicionados na ferramenta Parsifal conforme pode ser observado na Figura 9.

A definição das strings de busca e dos critérios de inclusão e exclusão, permitiu estabelecer um protocolo que serviu como base para a condução da RM apresentada na seção seguinte.

3.2 Condução

A condução de uma RM envolve a execução prática do plano estabelecido durante a fase de planejamento, seguindo o protocolo previamente definido.

Desta forma, primeiramente foram realizadas as buscas por trabalhos primários, utilizando a string de busca SB1 nas bases de dados escolhidas dentro do período entre 2015 e 2024. Os resultados apresentados para estas buscas são apresentados na Tabela 6, onde se pode verificar o número de trabalhos retornados para cada base de dados, somando um total de 385 trabalhos.

As etapas seguintes do processo consistem na extração de dados dos trabalhos encontrados por meio da string de busca SB1. Estas etapas podem se tornar complexas e suscetíveis a erros quando conduzidas manualmente devido a grande quantidade de dados. Por este motivo, torna-se necessária a utilização de ferramentas que permitam organizar, gerenciar, validar e analisar os dados extraídos. Por tanto, com a finalidade de facilitar estas tarefas, os trabalhos encontrados por meio da busca foram importados para o Parsifal, onde é possível gerenciar a seleção e aplicar os critérios de inclusão e exclusão definidos na etapa de planejamento. Esta funcionalidade é apresentada na Figura 10, onde é possível observar a lista com os trabalhos importados.

Tabela 6: Resultado da busca utilizando a string de busca SB1.

Base	**Resultados**
IEEE Xplore	29
ACM Digital Library	352
PubMed	5
Scopus	35
ProQuest	11
Total	**432**

Fonte: Autor

Após a importação dos trabalhos para o Parsifal foi realizada a seleção por meio da leitura dos títulos e resumos. Esta seleção resultou em:

- 14 trabalhos duplicados;
- 381 trabalhos rejeitados
- 37 trabalhos selecionados

Figura 10: Seleção de estudos na ferramenta Parsifal.

Find Duplicates | Export Articles

Action: Select... Go 0 of 375 selected

Show: All Accepted Rejected Unclassified Duplicated

Bibtex Key	Title	Author	Journal	Year	Added by	Added at	Status
	The Impact of High-Density Urban Wind Environments on the Distribution of COVID-19 Based on Machine Learning: A Case Study of Macau	Zheng, Liang and Chen, Yile and Yan, Lina and Zheng, Jianyi	Buildings	2023	marcelosf	04 Nov 2024 02:19:51	Rejected
	Soft computing of a medically important arthropod vector with autoregressive recurrent and focused time delay artificial neural networks	Damos, Petros and Tuells, José and Caballero, Pablo	Insects	2021	marcelosf	04 Nov 2024 02:19:51	Rejected
	Assessing the effect of Aedes (Stegomyia) aegypti (Linnaeus, 1762) control based on machine learning for predicting the spatiotemporal distribution of eggs in ovitraps	Piovezan, Rafael and de Azevedo, Thiago Salomão and Faria, Euler and Veroneze, Rosana and Von Zuben, Claudio José and Von Zuben, Fernando José and Sallum, Maria Anice Mureb	Dialogues in Health	2022	marcelosf	04 Nov 2024 02:19:51	Accepted
	Impacts of socioeconomic and environmental factors on neoplasms incidence rates using machine learning and GIS: a cross-sectional study in Iran	Rafiee, Mohammad and Jahangiri-rad, Mahsa and Mohseni-Bandpei, Anoushiravan and Razmi, Elham	Scientific Reports	2024	marcelosf	04 Nov 2024 02:19:51	Rejected
	Exploring the impact of socioeconomic and natural factors on pulmonary tuberculosis incidence in China (2013–2019) using explainable machine	Tao, Yiwen and Zhao, Jiaxin and Cui, Hao and Liu, Lili and He, Long	Acta Tropica	2024	marcelosf	04 Nov 2024 02:19:51	Rejected

Fonte: Autor.

Onde os trabalhos duplicados foram identificados de forma automática pelo Parsifal, que sugeriu a correção por meio da escolha de um dos trabalhos dentro da duplicidade. Já com relação aos trabalhos rejeitados, 381 não apresentaram alguma abordagem de ML que auxiliasse na correlação entre mudanças climáticas e doenças relacionadas ao clima. Deste modo, esta etapa resultou na seleção dos 37 trabalhos restantes, como poder ser visto na Tabela 7, onde são apresentadas a quantidade de trabalhos selecionados para cada base de dados, além do percentual de seleção.

Na última etapa de seleção, foram aplicados os critérios de inclusão e exclusão. Durante esta etapa, foram excluídos 6 trabalhos por não possuírem acesso disponível, diminuindo para 31 a quantidade de trabalhos selecionados. A Tabela 8 apresenta de forma sumarizada todo o processo de seleção dos trabalhos de acordo com cada base de dados, enquanto na Figura 11 é apresentado o fluxo de todo o processo de seleção dos trabalhos.

Tabela 7: Trabalhos selecionados após a leitura do título e resumo.

Base de dados	**Número de trabalhos selecionados**	**Percentual de seleção (%)**
IEEE Xplore	14	35,91
ACM Digital Library	6	15,38
PubMed	1	2,56
Scopus	15	38,46
ProQuest	3	7,69
Total	**39**	**100**

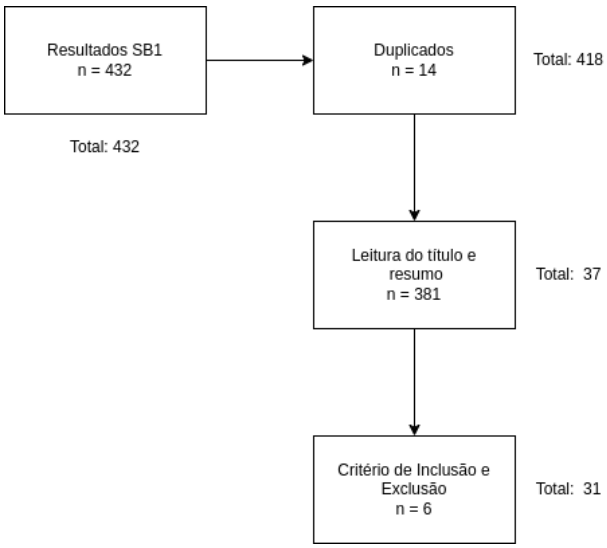
Fonte: Autor.

Tabela 8: Número de trabalhos selecionados em cada etapa de seleção.

Base de dados	**Busca (SB1)**	**Remoção de duplicidade**	**Leitura do título e resumo**	**Critérios de seleção**
IEEE Xplore	29	29	14	14
ACM Digital Library	352	351	6	2
PubMed	5	1	1	0
Scopus	35	33	13	12
ProQuest	11	4	3	3
Total	**432**	**418**	**37**	**31**

Fonte: Autor.

Figura 11: Processo de seleção dos trabalhos.



Fonte: Autor.

4 RESULTADOS

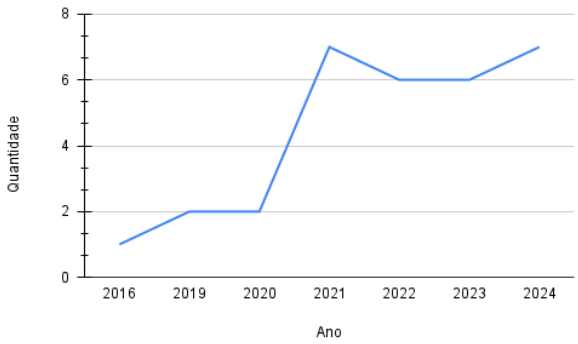
Este capítulo tem por objetivo apresentar a análise dos resultados obtidos por meio da execução do protocolo anteriormente estabelecido, contemplando todos os trabalhos selecionados.

4.1 Sumarização dos resultados

A RM conduzida neste trabalho identificou 31 estudos após a aplicação dos critérios de inclusão e exclusão, cobrindo um intervalo temporal de 8 anos, conforme apresentado na Figura 12. Os estudos selecionados foram obtidos das bases acadêmicas: Scopus, IEEE Xplore, ProQuest, PubMed e ACM Digital Library, e mostraram uma concentração significativa de publicações nos últimos 8 anos, refletindo o crescimento do interesse no tema.

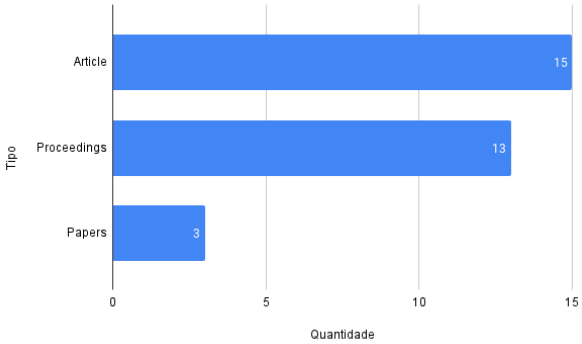
Os dados coletados mostram que a maioria dos estudos foram publicados entre os anos de 2020 e 2024, com um pico em 2024, evidenciando o aumento do uso de ML no contexto de mudanças climáticas, como pode ser visto na Figura 12. Já com relação as fontes utilizadas, a Tabela 8 mostra que o IEEE Xplore e Scopus se destacaram com a

Figura 12: Número de trabalhos por ano



Fonte: Autor.

Figura 13: Tipos de publicação



Fonte: Autor.

maior quantidade de estudos selecionados.

Quanto ao tipo de publicação, os estudos variam entre artigos, proceedings e papers sendo que 15 são artigos, 13 são proceedings e os outros 3 são papers, como mostra a Figura 13

Além disso, 289 autores contribuíram para a publicação destes 31 estudos selecionados, sendo que alguns participaram de mais de uma publicação. A relação entre autores e quantidade de publicações é apresentada na Tabela 9. Estes estudos abrangeram a aplicação de ML nas seguintes áreas principais:

- **Saúde pública:** Focando em doenças infecciosas, cardiovasculares e respiratórias;
- **Gestão de recursos da saúde:** Promovendo alocação eficiente de recursos com base na previsão de surtos.

A relação de estudos selecionados e seus respectivos autores é apresentada na Tabela 10

Tabela 9: Número de artigos por autor.

Autor	**Número de artigos**
Ahmed	2
Alex	2
Alireza	2
Badirzadeh	2
Dong	2
Gholami	2
Hamed	3
Hossein	2
Jiang	3
Kumar	2
Mitra	2
Mohammadi	4
Nguyen	3
Zahra	2
Zhang	2
Asadgol	2

Fonte: Autor

Tabela 10: Estudos selecionados e seus respectivos autores.

Título	**Autores**
Predicting the impact of climate change on the re-emergence of malaria cases in China using LSTMSeq2Seq deep learning model: a modelling and prediction analysis study	Kamana et al. (2022)
Identifying and predicting climate change impact on vector-borne disease using machine learning: Case study of Plasmodium falciparum from Africa	Singh e Saran (2024)
Forecasting habitat changes of Vimba persa (Pallas, 1814) under climate change using machine learning techniques in the southern Caspian Sea basin	Moëzzi et al. (2022)
Analysis of Correlation between Climate Change and Human Health Based on a Machine Learning Approach	Pizzulli et al. (2021a)
Impact of Climate Change and Air Pollution Forecasting Using Machine Learning Techniques in Bishkek	Isaev et al. (2022)
Prediction of Oncomelania hupensis distribution in association with climate change using machine learning models	Xu et al. (2023)
Toward Health-Centered Monitoring framework promoting Well-Being and Nature-based Solutions	Zafeiropoulos et al. (2024)
Tackling Climate Change with Machine Learning	Rolnick et al. (2022)
Pollen Allergy Prediction: A Review of the Time-Frame-Specific Forecast Models	Sharma e Kumar (2023)
Analysis and Prediction of Climate Change in Post-Covid19 India	Debdas et al. (2020)
Understanding the Dynamics of Dengue in Bangladesh: EDA, Climate Correlation, and Predictive Modeling	Meem et al. (2023)
Dengue Fever: From Extreme Climates to Outbreak Prediction	Mai et al. (2022)
Diarrhoea incidence prediction using climate data: Machine Learning approaches	Do et al. (2022a)

Predicting Disease Outbreaks with Climate Data	Abdullahi e Nitschke (2021)
Analyzing Weather Patterns and Mosquito Data to Predict West Nile Virus Outbreaks in Chicago	Bachal et al. (2023)
A Machine Learning Based Analysis Between Climate Change and Human Health: A Correlational Study	Fahim et al. (2022b)
Using Machine Learning to Analyze and Predict the Relations Between Cardiovascular Disease Incidence, Extreme Temperature and Air Pollution	Lin et al. (2021a)
Cluster based regression model on dengue incidence using dual climate variables	Mathulamuthu et al. (2016)
AI-based prediction for early detection of Tuberculosis in India based on environmental factors	Giri et al. (2020a)
Applications of artificial intelligence in predicting dengue outbreaks in the face of climate change: a case study along coastal India	Tallam e Quang (2023)
Modeling climate change impacts on vector-borne disease using machine learning models: Case study of Visceral leishmaniasis (Kala-azar) from Indian state of Bihar	Kumar et al. (2024)
Predictive modeling of suitable habitats for cinnamomum camphora (L.) presl using maxent model under climate change in China	Zhang et al. (2019)
Using artificial neural networks to model the impacts of climate change on dust phenomenon in the Zanjan region, north-west Iran	Moghanlo et al. (2021)
Sound-based Mosquito Classification via Featurization and Machine Learning	Gupta e Gadre (2024)
Simulation of the potential impact of climate change on malaria incidence using artificial neural networks (ANNs)	Asadgol et al. (2023)

Prediction of Dengue Fever Outbreak Based on Climate and Demographic Variables Using Extreme Gradient Boosting and Rule-Based Classification	Anggraeni et al. (2021)
Geospatial Analysis of Extreme Temperature Impacts in Agricultural Systems Using Machine Learning	Wheeler et al. (2024)
The effect of climate change on cholera disease: The road ahead using artificial neural network	Asadgol et al. (2019)
Prediction of malaria incidence using climate variability and machine learning	Nkiruka et al. (2021)
Climate change characteristics and population health impact factors using deep neural network and hyperautomation mechanism	Shao e Zhang (2024)
Quantitative Impact Analysis of Climate Change on Residents' Health Conditions with Improving Eco-Efficiency in China: A Machine Learning Perspective	Wang et al. (2021)

Fonte: Autor.

4.2 Algoritmos

Por meio dos dados coletados, foi possível obter os diferentes tipos de algoritmos de ML abordados nos estudos encontrados. A Figura 15, apresenta estes algoritmos e o número de vezes que foram utilizados pelos estudos selecionados. Ainda na Figura 15, é possível notar a prevalência na utilização de Redes Neurais Artificiais, sendo abordado em 15 estudos, e em seguida, os algoritmos Support Vector Machine Extreme Gradient Boosting (XGBoost) e Random Forest.

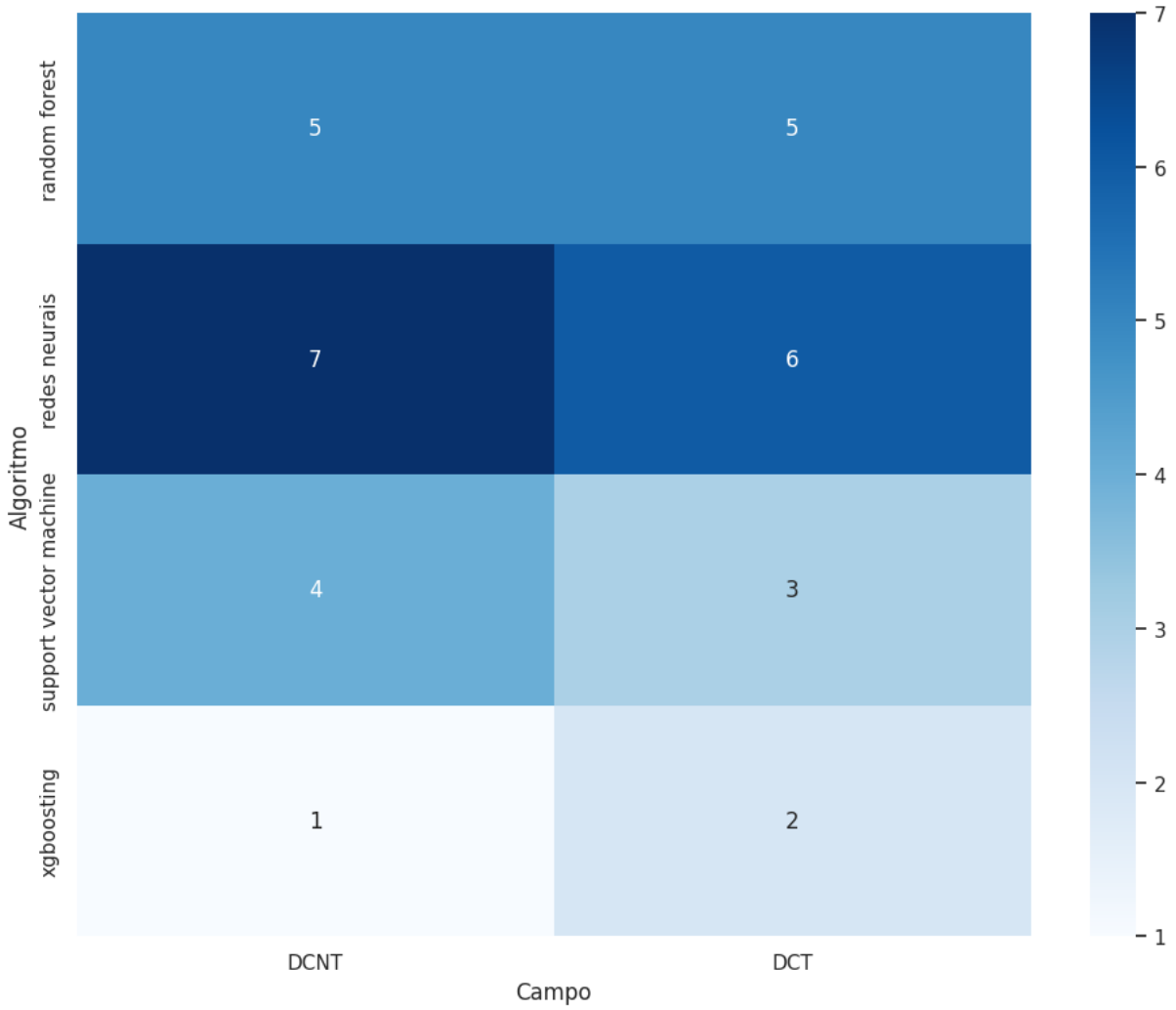
Quanto ao campo de aplicação, foi possível criar as seguintes classificações:

- **Doenças crônicas transmissíveis (DCT):** se refere á aplicação de ML para a predição da incidência de doenças transmissíveis causadas por vetores;
- **Doenças crônicas não transmissíveis (DCNT):** se refere a utilização de ML na predição de doenças respiratórias, cardiovasculares e mentais.

Esta classificação é representada pela Figura 14, onde é possível notar que, dentro do conjunto de estudos selecionados, o campo com maior utilização de ML foi para predição de DCNT com a utilização de Redes Neurais Artificiais, seguido pela utilização do algoritmo Random Forest.

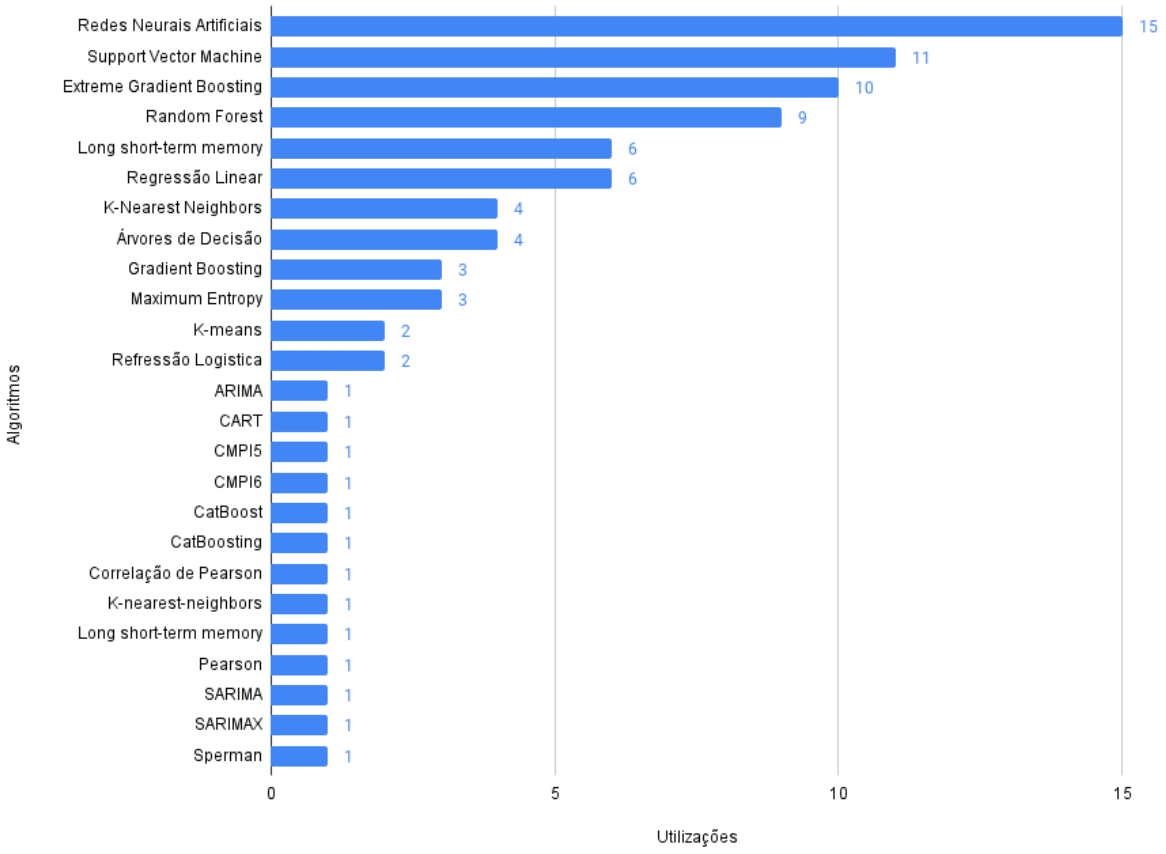
A relação entre os algoritmos mais abordados e os trabalhos que os utilizaram é apresentada na Tabela 11

Figura 14: Utilização dos algoritmos de acordo com o campo de estudo



Fonte: Autor.

Figura 15: Utilização de cada algoritmo.



Fonte: Autor.

Tabela 11: Relação entre estudos selecionados e algoritmos mais utilizados.

Algoritmo	**Estudo**
Redes Neurais Artificiais	Asadgol et al. (2019), Xu et al. (2023), Moghanlo et al. (2021), Do et al. (2022b), Rolnick et al. (2022), Kumar et al. (2024), Gupta e Gadre (2024), Giri et al. (2020b), Sharma e Kumar (2023), Abdullahi e Nitschke (2021), Isaev et al. (2022), Pizzulli et al. (2021b), Asadgol et al. (2023), Singh e Saran (2024), Shao e Zhang (2024)
Support Vector Machine	Xu et al. (2023), Rolnick et al. (2022), Tallam e Quang (2023), Kumar et al. (2024), Sharma e Kumar (2023), Abdullahi e Nitschke (2021), Singh e Saran (2024)
Extreme Gradient Boosting	Xu et al. (2023), Anggraeni et al. (2021), Bachal et al. (2023), Giri et al. (2020a), Isaev et al. (2022), Nkiruka et al. (2021)
Random Forest	Xu et al. (2023), Rolnick et al. (2022), Tallam e Quang (2023), Giri et al. (2020b), Sharma e Kumar (2023), Isaev et al. (2022), Nkiruka et al. (2021)

Fonte: Autor.

4.3 Capacidade de detecção e previsão

A análise da revisão de mapeamento baseada nos 33 estudos selecionados revelou a capacidade das técnicas de Machine Learning (ML) em detectar correlações entre variáveis climáticas, como temperatura, umidade e padrões de precipitação, e a incidência de doenças sensíveis ao clima, incluindo malária, dengue e doenças respiratórias. Os estudos revisados destacam o uso de modelos preditivos, análises de correlação e aprendizado profundo para quantificar os impactos dessas variáveis ambientais na saúde pública.

É o caso de Giri et al. (2020b), que demonstra como fatores climáticos, como umidade relativa e temperatura média, podem ser utilizados em modelos de regressão para prever surtos de tuberculose, com correlações observadas entre níveis elevados de poluição e a prevalência da doença. Similarmente, o estudo de Fahim et al. (2022a) utiliza algoritmos como Pearson e Spearman para evidenciar uma correlação de 98% entre dióxido de carbono e doenças cardiovasculares, reforçando o impacto das mudanças climáticas na saúde humana.

Estudos relacionados à malária, como Kimuyu (2021), indicam que mudanças nos padrões de precipitação estão diretamente relacionadas à proliferação de vetores em áreas anteriormente não endêmicas, destacando a necessidade de políticas de mitigação informadas por dados. Além disso, o estudo de Lin et al. (2021b) reforça a influência de fatores climáticos e poluentes, como o dióxido de enxofre, em condições crônicas respiratórias e cardiovasculares, utilizando redes neurais artificiais para prever tendências futuras.

No campo das doenças respiratórias, o estudo de D'Amato et al. (2016) destaca que variações na qualidade do ar, como concentrações de partículas finas (PM2.5), contribuem significativamente para o aumento de alergias respiratórias e doenças pulmonares em áreas urbanas densamente povoadas.

Esses resultados evidenciam que as técnicas de ML são ferramentas úteis na integração de grandes volumes de dados ambientais e epidemiológicos, fornecendo insights preditivos que podem apoiar a formulação de políticas públicas mais eficazes. No entanto, desafios como a qualidade dos dados e a complexidade computacional dos modelos requerem maior atenção para maximizar os benefícios dessas tecnologias na saúde pública.

4.4 Políticas de saúde pública

Os estudos analisados exploram como as técnicas de ML podem ser aplicadas para compreender e mitigar os impactos das mudanças climáticas dentro do contexto de saúde pública. Essas abordagens demonstram o potencial dos modelos preditivos e análises correlacionais para auxiliar em decisões políticas e estratégicas de intervenção, considerando dados ambientais, epidemiológicos e sociais.

No estudo de Pizzulli et al. (2021b), identificou-se que as mudanças de temperatura estão correlacionadas com o aumento de doenças respiratórias e do sistema nervoso, enquanto doenças digestivas demonstraram relação fraca com variáveis climáticas. Essa análise sugere que abordagens regionais mais detalhadas podem aprimorar a robustez dos modelos, fornecendo subsídios para políticas direcionadas.

Já o estudo de, Kimuyu (2021), utilizou modelos de ML para prever a distribuição de vetores de malária em regiões não endêmicas. Este trabalho destaca a importância de incluir comunidades locais no planejamento e na mitigação de riscos em hotspots potenciais, exemplificando como os resultados da utilização de ML podem fortalecer estratégias de controle de vetores e intervenções equitativas.

Estudos focados em doenças específicas, como dengue e malária, também reforçam o valor das técnicas de ML. O estudo Tian et al. (2024) demonstra como modelos como o XGBoost realizaram a correlação entre variáveis climáticas e epidemiológicas, oferecendo insights para estratégias de controle vetorial. Enquanto os modelos de aprendizado profundo, como o LSTMSeq2Seq, foram úteis na previsão da reemergência de malária na China, indicando a necessidade de intervenções preventivas baseadas em dados climáticos.

Além disso, abordagens como as de Lin et al. (2021b), mostram como múltiplos fatores ambientais podem ser correlacionados com doenças cardiovasculares. Esses estudos sugerem que a inclusão de variáveis contextuais e o uso de dados em tempo real melhoram a precisão preditiva e apoiam políticas de saúde pública mais ágeis e informadas.

Esses estudos evidenciam como a utilização de ML é capaz de transformar grandes volumes de dados em insights importantes para a tomada de decisões, beneficiando políticas de saúde pública. Apesar das limitações associadas à qualidade e disponibilidade de dados, a interdisciplinaridade e o avanço tecnológico continuam a expandir a aplicabilidade dessas abordagens, permitindo intervenções mais eficazes frente aos desafios das mudanças climáticas na saúde pública.

5 CONCLUSÃO

Os estudos analisados evidenciam o crescente potencial da utilização de ML como ferramenta estratégica para compreender e mitigar os impactos das mudanças climáticas sobre a saúde pública. As aplicações dessas técnicas abrangem desde a previsão de doenças relacionadas ao clima, como dengue, malária e doenças respiratórias, até a avaliação de impactos em saúde mental e segurança alimentar. Os modelos de ML mostraram-se eficazes na análise de grandes volumes de dados complexos, permitindo identificar padrões e correlações que seriam difíceis de observar por métodos tradicionais.

Estas técnicas foram abordaddos em 31 estudos selecionados por meio da Revisão de Mapeamento, onde os autores mais recorrentes foram: Hamed, Jiang, Mohammadi e Nguyen. A Revisão de Mapeamento permitiu ainda classificar os algoritmos de ML quanto o campo de aplicação da seguinte forma: Doenças crônicas transmissíveis (DCT) e Doenças crônicas não transmissíveis (DCNT), onde as Redes Neurais Artificiais foram mais utilizadas dentro da classificação DCNT

Este estudo também evidenciou os algoritmos mais utilizados considerando os 31 trabalhos selecionados, sendo eles: Redes Neurais Artificiais, Support Vector Machine, Extreme Gradient Boosting e Random Forest.

Contudo, os desafios associados à qualidade e disponibilidade dos dados, à complexidade dos modelos e à necessidade de interpretações confiáveis para a tomada de decisão permanecem como barreiras. A integração de dados contextuais, como variáveis socioeconômicas e regionais, e a colaboração interdisciplinar são passos importantes para aprimorar a aplicabilidade prática das abordagens desenvolvidas. Além disso, políticas públicas baseadas em evidências geradas por ML podem desempenhar um papel essencial na alocação eficiente de recursos e no planejamento de estratégias adaptativas.

Por fim, o avanço contínuo das tecnologias de ML, aliado a esforços conjuntos entre pesquisadores, gestores públicos e comunidades, tem o potencial de transformar os desafios impostos pelas mudanças climáticas em oportunidades para fortalecer a saúde pública

global, promovendo resiliência e equidade no enfrentamento dessas adversidades.

REFERÊNCIAS

ABDULLAHI, T.; NITSCHKE, G. Predicting disease outbreaks with climate data. In: **2021 IEEE Congress on Evolutionary Computation (CEC)**. [S.l.: s.n.], 2021. p. 989–996.

Prediction of Dengue Fever Outbreak Based on Climate and Demographic Variables Using Extreme Gradient Boosting and Rule-Based Classification. 1-8 p.

ANSAH, E. W.; AMOADU, M.; OBENG, P.; SARFO, J. O. Health systems response to climate change adaptation: a scoping review of global evidence. **BMC Public Health**, v. 24, n. 1, p. 2015, Jul 2024. ISSN 1471-2458. Disponível em: <https://doi.org/10.1186/s12889-024-19459-w>.

APETORGBOR, M.; WANKHEDE, V.; DAKHARE, K.; VERMA, P.; RAUT, Y.; TRIPATHI, C. Application of machine learning for detection and prediction of malaria: A review. In: **2024 3rd International Conference for Innovation in Technology (INOCON)**. [S.l.: s.n.], 2024. p. 1–5.

ARTHUR, D.; VASSILVITSKII, S. k-means++: the advantages of careful seeding. In: **Proceedings of the Eighteenth Annual ACM-SIAM Symposium on Discrete Algorithms**. USA: Society for Industrial and Applied Mathematics, 2007. (SODA '07), p. 1027–1035. ISBN 9780898716245.

ASADGOL, Z.; BADIRZADEH, A.; MIRAHMADI, H.; SAFARI, H.; MOHAMMADI, H.; GHOLAMI, M. Simulation of the potential impact of climate change on malaria incidence using artificial neural networks (anns). **Environmental Science and Pollution Research**, v. 30, n. 30, p. 75349–75368, Jun 2023. ISSN 1614-7499. Disponível em: <https://doi.org/10.1007/s11356-023-27374-7>.

ASADGOL, Z.; MOHAMMADI, H.; KERMANI, M.; BADIRZADEH, A.; GHOLAMI, M. The effect of climate change on cholera disease: The road ahead using artificial neural network. **PLoS One**, United States, v. 14, n. 11, p. e0224813, nov. 2019.

BACHAL, T.; KUMAR, A.; SHETTY, A.; PAWAR, R. Analyzing weather patterns and mosquito data to predict west nile virus outbreaks in chicago. In: **2023 7th International Conference On Computing, Communication, Control And Automation (ICCUBEA)**. [S.l.: s.n.], 2023. p. 1–5.

BERRANG-FORD, L.; SIETSMA, A. J.; CALLAGHAN, M.; MINX, J. C.; SCHEELBEEK, P. F. D.; HADDAWAY, N. R.; HAINES, A.; DANGOUR, A. D. Systematic mapping of global research on climate and health: a machine learning review. **The Lancet Planetary Health**, Elsevier, v. 5, n. 8, p. e514–e525, Aug 2021. ISSN 2542-5196. Disponível em: <https://doi.org/10.1016/S2542-5196(21)00179-0>.

BERRY, H. L.; BOWEN, K.; KJELLSTROM, T. Climate change and mental health: a causal pathways framework. **Int J Public Health**, Switzerland, v. 55, n. 2, p. 123–132, dez. 2009.

BRAUNWALD, E. Cardiovascular effects of climate change. **European Heart Journal**, v. 45, n. 33, p. 3006–3008, 07 2024. ISSN 0195-668X. Disponível em: <https://doi.org/10.1093/eurheartj/ehae401>.

BREIMAN, L. Bagging predictors. **Machine Learning**, Springer, v. 24, n. 2, p. 123–140, 1996.

_____. Random forests. **Machine Learning**, Springer, v. 45, n. 1, p. 5–32, 2001.

BREIMAN, L.; FRIEDMAN, J. H.; OLSHEN, R. A.; STONE, C. J. **Classification and regression trees**. [S.l.]: Wadsworth International Group, 1984.

CALVIN, K.; DASGUPTA, D.; KRINNER, G.; MUKHERJI, A.; THORNE, P. W.; TRISOS, C.; ROMERO, J.; ALDUNCE, P.; BARRETT, K.; BLANCO, G.; CHEUNG, W. W.; CONNORS, S.; DENTON, F.; DIONGUE-NIANG, A.; DODMAN, D.; GARSCHAGEN, M.; GEDEN, O.; HAYWARD, B.; JONES, C.; JOTZO, F.; KRUG, T.; LASCO, R.; LEE, Y.-Y.; MASSON-DELMOTTE, V.; MEINSHAUSEN, M.; MINTENBECK, K.; MOKSSIT, A.; OTTO, F. E.; PATHAK, M.; PIRANI, A.; POLOCZANSKA, E.; PöRTNER, H.-O.; REVI, A.; ROBERTS, D. C.; ROY, J.; RUANE, A. C.; SKEA, J.; SHUKLA, P. R.; SLADE, R.; SLANGEN, A.; SOKONA, Y.; SöRENSSON, A. A.; TIGNOR, M.; VUUREN, D. van; WEI, Y.-M.; WINKLER, H.; ZHAI, P.; ZOMMERS, Z.; HOURCADE, J.-C.; JOHNSON, F. X.; PACHAURI, S.; SIMPSON, N. P.; SINGH, C.; THOMAS, A.; TOTIN, E.; ALEGRíA, A.; ARMOUR, K.; BEDNAR-FRIEDL, B.; BLOK, K.; CISSé, G.; DENTENER, F.; ERIKSEN, S.; FISCHER, E.; GARNER, G.; GUIVARCH, C.; HAASNOOT, M.; HANSEN, G.; HAUSER, M.; HAWKINS, E.; HERMANS, T.; KOPP, R.; LEPRINCE-RINGUET, N.; LEWIS, J.; LEY, D.; LUDDEN, C.; NIAMIR, L.; NICHOLLS, Z.; SOME, S.; SZOPA, S.; TREWIN, B.; WIJST, K.-I. van der; WINTER, G.; WITTING, M.; BIRT, A.; HA, M. **IPCC, 2023: Climate Change 2023: Synthesis Report. Contribution of Working Groups I, II and III to the Sixth Assessment Report of the Intergovernmental Panel on Climate Change [Core Writing Team, H. Lee and J. Romero (eds.)]. IPCC, Geneva, Switzerland.** [s.n.], 2023. Disponível em: <http://dx.doi.org/10.59327/ipcc/ar6-9789291691647>.

CAMPBELL, A. M.; RACAULT, M.-F.; GOULT, S.; LAURENSON, A. Cholera risk: A machine learning approach applied to essential climate variables. **International Journal of Environmental Research and Public Health**, v. 17, 2020. ISSN 1660-4601. Este artigo faz uma exploração a respeito do potencialda aplicação de ML na previsão do risco de colera na costa da India, considerando variáveis climáticas atmosféricas, terrestres. Disponível em: <https://www.mdpi.com/1660-4601/17/24/9378>.

CHATTERJEE, S.; HADI, A. S. Influential Observations, High Leverage Points, and Outliers in Linear Regression. **Statistical Science**, Institute of Mathematical Statistics, v. 1, n. 3, p. 379 – 393, 1986. Disponível em: <https://doi.org/10.1214/ss/1177013622>.

CLAYTON, S.; MANNING, C.; HODGE, C. **Beyond storms and droughts: The psychological impacts of climate change**. [S.l.]: American Psychological Association, 2017.

CONFALONIERI, U.; MENNE, B.; AKHTAR, R.; EBI, K. L.; HAUENGUE, M.; KOVATS, S. R.; WOODWARD, A. et al. Human health. In: PARRY, M. L.; CANZIANI, O. F.; PALUTIKOF, J. P.; LINDEN, P. J. van der; HANSON, C. E. (Ed.). **Climate Change: Impacts, Adaptation and Vulnerability. Contribution of Working Group II to the Fourth Assessment Report of the Intergovernmental Panel on Climate Change**. Cambridge, UK: Cambridge University Press, 2009. p. 391–431.

CUI, H.; HU, L.; CHI, L. Advances in computer-aided medical image processing. **Applied Sciences**, v. 13, n. 12, 2023. ISSN 2076-3417. Disponível em: <https://www.mdpi.com/2076-3417/13/12/7079>.

CUTLER, D. R.; JR., T. C. E.; BEARD, K. H.; CUTLER, A.; HESS, K. T.; GIBSON, J.; LAWLER, J. J. Random forests for classification in ecology. **Ecology**, v. 88, n. 11, p. 2783–2792, 2007. Disponível em: <https://esajournals.onlinelibrary.wiley.com/doi/abs/10.1890/07-0539.1>.

D'AMATO, G.; PAWANKAR, R.; VITALE, C.; LANZA, M.; MOLINO, A.; STANZIOLA, A.; SANDUZZI, A.; VATRELLA, A.; D'AMATO, M. Climate change and air pollution: Effects on respiratory allergy. **Allergy Asthma Immunol Res**, Korea (South), v. 8, n. 5, p. 391–395, set. 2016.

DEBDAS, S.; ROY, K.; SAHA, A.; KUNDU, S.; DASRAY, S.; CHOUHAN, S. Analysis and prediction of climate change in post-covid19 india. In: **2020 4th International Conference on Electronics, Communication and Aerospace Technology (ICECA)**. [S.l.: s.n.], 2020. p. 1530–1536.

DO, T. D.; MULHALL, J.; NGUYEN, T. D.; NGUYEN, Q. T.; PHAN, D.; MAI, S. T. Diarrhoea incidence prediction using climate data: Machine learning approaches. In: **2022 RIVF International Conference on Computing and Communication Technologies (RIVF)**. [S.l.: s.n.], 2022. p. 1–6.

_____. Diarrhoea incidence prediction using climate data: Machine learning approaches. In: **2022 RIVF International Conference on Computing and Communication Technologies (RIVF)**. [S.l.: s.n.], 2022. p. 1–6.

EBI, K.; MCGREGOR, G. Climate change, tropospheric ozone and particulate matter, and health impacts. **Ciência Saúde Coletiva**, ABRASCO - Associação Brasileira de Saúde Coletiva, v. 14, n. 6, p. 2281–2293, Dec 2009. ISSN 1413-8123. Disponível em: <https://doi.org/10.1590/S1413-81232009000600037>.

EDELSON, P. J.; HAROLD, R.; ACKELSBERG, J.; DUCHIN, J. S.; LAWRENCE, S. J.; MANABE, Y. C.; ZAHN, M.; LAROCQUE, R. C. Climate Change and the Epidemiology of Infectious Diseases in the United States. **Clinical Infectious Diseases**, v. 76, n. 5, p. 950–956, 09 2022. ISSN 1058-4838. Disponível em: <https://doi.org/10.1093/cid/ciac697>.

ENGELEN, J. E. van; HOOS, H. H. A survey on semi-supervised learning. **Machine Learning**, v. 109, n. 2, p. 373–440, Feb 2020. ISSN 1573-0565. Disponível em: <https://doi.org/10.1007/s10994-019-05855-6>.

FAHIM, M.; UDDIN, M. E.; AHMED, R.; ISLAM, M. R.; AHMED, N. A machine learning based analysis between climate change and human health: A correlational study. In: **2022 International Conference on Computer and Applications (ICCA)**. [S.l.: s.n.], 2022. p. 1–6.

_____. A machine learning based analysis between climate change and human health: A correlational study. In: **2022 International Conference on Computer and Applications (ICCA)**. [S.l.: s.n.], 2022. p. 1–6.

FRENK, J. The global health system: strengthening national health systems as the next step for global progress. **PLoS Med**, United States, v. 7, n. 1, p. e1000089, jan. 2010.

GIRI, N.; JOSEPH, R.; CHAVAN, S.; HEDA, R.; ISRANI, R.; SETHIYA, R. Ai-based prediction for early detection of tuberculosis in india based on environmental factors. In: **2020 19th IEEE International Conference on Machine Learning and Applications (ICMLA)**. [S.l.: s.n.], 2020. p. 278–285.

_____. Ai-based prediction for early detection of tuberculosis in india based on environmental factors. In: **2020 19th IEEE International Conference on Machine Learning and Applications (ICMLA)**. [S.l.: s.n.], 2020. p. 278–285.

GUPTA, D.; GADRE, A. Sound-based mosquito classification via featurization and machine learning. **2024 IEEE 7th International Conference on Big Data and Artificial Intelligence (BDAI)**, p. 1–6, 2024. Disponível em: <https://api.semanticscholar.org/CorpusID:273027890>.

HALES, S.; WET, N. de; MAINDONALD, J.; WOODWARD, A. Potential effect of population and climate changes on global distribution of dengue fever: an empirical model. **The Lancet**, Elsevier, v. 360, n. 9336, p. 830–834, Sep 2002. ISSN 0140-6736. Disponível em: <https://doi.org/10.1016/S0140-6736(02)09964-6>.

HARTIGAN, J. A.; WONG, M. A. Algorithm AS 136: A k-means clustering algorithm. **Journal of the Royal Statistical Society. Series C (Applied Statistics)**, Wiley, v. 28, n. 1, p. 100–108, 1979.

HOCHREITER, S.; SCHMIDHUBER, J. Long Short-Term Memory. **Neural Computation**, v. 9, n. 8, p. 1735–1780, 11 1997. ISSN 0899-7667. Disponível em: <https://doi.org/10.1162/neco.1997.9.8.1735>.

ISAEV, E.; AJIKEEV, B.; SHAMYRKANOV, U.; KALNUR, K.-u.; MAISALBEK, K.; SIDLE, R. C. Impact of climate change and air pollution forecasting using machine learning techniques in bishkek. **Aerosol and Air Quality Research**, Taiwan Association for Aerosol Research, v. 22, n. 3, p. 210336, 2022. ISSN 2071-1409. Disponível em: <http://dx.doi.org/10.4209/aaqr.210336>.

KAMANA, E.; ZHAO, J.; BAI, D. Predicting the impact of climate change on the re-emergence of malaria cases in china using LSTMSeq2Seq deep learning model: a modelling and prediction analysis study. **BMJ Open**, England, v. 12, n. 3, p. e053922, mar. 2022.

KAZI, D. S.; KATZNELSON, E.; LIU, C.-L.; AL-ROUB, N. M.; CHAUDHARY, R. S.; YOUNG, D. E.; MCNICHOL, M.; MICKLEY, L. J.; KRAMER, D. B.; CASCIO, W. E.; BERNSTEIN, A. S.; RICE, M. B. Climate change and cardiovascular health: A systematic review. **JAMA Cardiol**, United States, v. 9, n. 8, p. 748–757, ago. 2024.

KIMUYU, J. S. Comparative spatial–temporal analysis and predictive modeling of climate change-induced malaria vectors' invasion in new hotspots in kenya. **SN Applied Sciences**, v. 3, n. 8, p. 741, Jul 2021. ISSN 2523-3971. Disponível em: <https://doi.org/10.1007/s42452-021-04722-1>.

KITCHENHAM, B. Procedures for performing systematic reviews. **Keele, UK, Keele Univ.**, v. 33, 08 2004.

KITCHENHAM, B. A.; BUDGEN, D.; Pearl Brereton, O. Using mapping studies as the basis for further research – a participant-observer case study. **Information and Software Technology**, v. 53, n. 6, p. 638–651, 2011. ISSN 0950-5849. Special Section: Best papers from the APSEC. Disponível em: <https://www.sciencedirect.com/science/article/pii/S0950584910002272>.

KITCHENHAM, B. A.; CHARTERS, S. **Guidelines for performing Systematic Literature Reviews in Software Engineering**. [S.l.], 2007. Disponível em: <https://www.elsevier.com/__data/promis_misc/525444systematicreviewsguide.pdf>.

KUMAR, S.; SRIVASTAVA, A.; MAITY, R. Modeling climate change impacts on vector-borne disease using machine learning models: Case study of visceral leishmaniasis (kala-azar) from indian state of bihar. **Expert Systems with Applications**, v. 237, p. 121490, 2024. ISSN 0957-4174. Disponível em: <https://www.sciencedirect.com/science/article/pii/S0957417423019929>.

LECUN, Y.; BENGIO, Y.; HINTON, G. Deep learning. **Nature**, v. 521, n. 7553, p. 436–444, May 2015. ISSN 1476-4687. Disponível em: <https://doi.org/10.1038/nature14539>.

LIAW, A.; WIENER, M. Classification and regression by randomforest. **R News**, The R Foundation, v. 2, n. 3, p. 18–22, 2002.

LIN, Y.-C.; TSAI, C.-H.; HSU, H.-T.; LIN, C.-H. Using machine learning to analyze and predict the relations between cardiovascular disease incidence, extreme temperature and air pollution. In: **2021 IEEE 3rd Eurasia Conference on Biomedical Engineering, Healthcare and Sustainability (ECBIOS)**. [S.l.: s.n.], 2021. p. 234–237.

_____. Using machine learning to analyze and predict the relations between cardiovascular disease incidence, extreme temperature and air pollution. In: **2021 IEEE 3rd Eurasia Conference on Biomedical Engineering, Healthcare and Sustainability (ECBIOS)**. [S.l.: s.n.], 2021. p. 234–237.

LLOYD, S. Least squares quantization in pcm. **IEEE Transactions on Information Theory**, v. 28, n. 2, p. 129–137, 1982.

LUBER, G.; PRUDENT, N. Climate change and human health. **Transactions of the American Clinical and Climatological Association**, v. 120, p. 113–117, 2009. Disponível em: <https://www.ncbi.nlm.nih.gov/pmc/articles/PMC2744549/>.

MACQUEEN, J. Some methods for classification and analysis of multivariate observations. In: **Proceedings of the Fifth Berkeley Symposium on Mathematical Statistics and Probability**. Berkeley, CA: University of California Press, 1967. v. 1, n. 14, p. 281–297.

MAI, S. T.; PHI, H. T.; ABUBAKAR, A.; KILPATRICK, P.; NGUYEN, H. Q. V.; VANDIERENDONCK, H. Dengue fever: From extreme climates to outbreak prediction. In: **2022 IEEE International Conference on Data Mining (ICDM)**. [S.l.: s.n.], 2022. p. 1083–1088.

MARSHALL, C.; BRERETON, P.; KITCHENHAM, B. Tools to support systematic reviews in software engineering: A feature analysis. In: . [S.l.: s.n.], 2014.

MATHULAMUTHU, S. S.; ASIRVADAM, V. S.; DASS, S. C.; GILL, B. S. Cluster based regression model on dengue incidence using dual climate variables. In: **2016 IEEE Conference on Systems, Process and Control (ICSPC)**. [S.l.: s.n.], 2016. p. 64–69.

MCCULLOCH, W. S.; PITTS, W. A logical calculus of the ideas immanent in nervous activity. **The bulletin of mathematical biophysics**, v. 5, n. 4, p. 115–133, Dec 1943. ISSN 1522-9602. Disponível em: <https://doi.org/10.1007/BF02478259>.

MEEM, S. M.; HOSSAIN, M. T.; CHOWDHURY, J. K.; MIAH, M. S. U.; MONIR, M. F. Understanding the dynamics of dengue in bangladesh: Eda, climate correlation, and predictive modeling. In: **TENCON 2023 - 2023 IEEE Region 10 Conference (TENCON)**. [S.l.: s.n.], 2023. p. 1309–1314.

MINSKY, M.; PAPERT, S. **Perceptrons: An Introduction to Computational Geometry**. Cambridge, MA, USA: MIT Press, 1969.

MOGHANLO, S.; ALAVINEJAD, M.; OSKOEI, V.; Najafi Saleh, H.; MOHAMMADI, A. A.; MOHAMMADI, H.; DERAKHSHANNEJAD, Z. Using artificial neural networks to model the impacts of climate change on dust phenomenon in the zanjan region, northwest iran. **Urban Climate**, v. 35, p. 100750, 2021. ISSN 2212-0955. Disponível em: <https://www.sciencedirect.com/science/article/pii/S2212095520306647>.

MOëZZI, F.; POORBAGHER, H.; EAGDERI, S.; FEGHHI, J.; DORMANN, C. F.; NERGI, S. K.; AMIRI, K. Modelling habitat preference of caspian kutum, rutilus kutum, using non-linear habitat suitability indices and generalized additive models. **Regional Studies in Marine Science**, v. 56, p. 102715, 2022. ISSN 2352-4855. Disponível em: <https://www.sciencedirect.com/science/article/pii/S2352485522003103>.

MUNN, Z.; PETERS, M. D. J.; STERN, C.; TUFANARU, C.; MCARTHUR, A.; AROMATARIS, E. Systematic review or scoping review? guidance for authors when choosing between a systematic or scoping review approach. **BMC Medical Research Methodology**, v. 18, n. 1, p. 143, Nov 2018. ISSN 1471-2288. Disponível em: <https://doi.org/10.1186/s12874-018-0611-x>.

N, T. R.; GUPTA, R. A survey on machine learning approaches and its techniques:. In: **2020 IEEE International Students' Conference on Electrical,Electronics and Computer Science (SCEECS)**. [S.l.: s.n.], 2020. p. 1–6.

NG, R. T.; HAN, J. CLARANS: A method for clustering objects for spatial data mining. **IEEE Transactions on Knowledge and Data Engineering**, IEEE, v. 14, n. 5, p. 1003–1016, 2002.

NICHOLS, G.; LAKE, I.; HEAVISIDE, C. Climate change and water-related infectious diseases. **Atmosphere**, MDPI AG, v. 9, n. 10, out. 2018. ISSN 2073-4433.

NKIRUKA, O.; PRASAD, R.; CLEMENT, O. Prediction of malaria incidence using climate variability and machine learning. **Informatics in Medicine Unlocked**, v. 22, p. 100508, 2021. ISSN 2352-9148. Disponível em: <https://www.sciencedirect.com/science/article/pii/S2352914820306596>.

Parsifal. **Parsifal: Systematic Literature Review Tool**. 2024. <https://parsif.al/about/>. Accessed: 2024-09-28.

PATEL, J.; SHAH, S.; THAKKAR, P.; KOTECHA, K. Predicting stock market index using fusion of machine learning techniques. **Expert Systems with Applications**, v. 42, 10 2014.

PATZ, J. A.; CAMPBELL-LENDRUM, D.; HOLLOWAY, T.; FOLEY, J. A. Impact of regional climate change on human health. **Nature**, v. 438, n. 7066, p. 310–317, Nov 2005. ISSN 1476-4687. Disponível em: <https://doi.org/10.1038/nature04188>.

PIZZULLI, V. A.; TELESCA, V.; COVATARIU, G. Analysis of correlation between climate change and human health based on a machine learning approach. **Healthcare (Basel)**, Switzerland, v. 9, n. 1, jan. 2021.

_____. Analysis of correlation between climate change and human health based on a machine learning approach. **Healthcare (Basel)**, Switzerland, v. 9, n. 1, jan. 2021.

POPE 3RD, C. A.; EZZATI, M.; DOCKERY, D. W. Fine-particulate air pollution and life expectancy in the united states. **N Engl J Med**, United States, v. 360, n. 4, p. 376–386, jan. 2009.

QUINLAN, J. R. **C4.5: Programs for machine learning**. [S.l.]: Morgan Kaufmann, 1993.

RACLOZ, V.; RAMSEY, R.; TONG, S.; HU, W. Surveillance of dengue fever virus: A review of epidemiological models and early warning systems. **PLoS Neglected Tropical Diseases**, v. 6, 2012. Disponível em: <https://api.semanticscholar.org/CorpusID:19024384>.

RAZDAN, T.; MITTAL, N. Climate change impact analysis on plantation. In: **2024 11th International Conference on Reliability, Infocom Technologies and Optimization (Trends and Future Directions) (ICRITO)**. [S.l.: s.n.], 2024. p. 1–5.

ROLNICK, D.; DONTI, P. L.; KAACK, L. H.; KOCHANSKI, K.; LACOSTE, A.; SANKARAN, K.; ROSS, A. S.; MILOJEVIC-DUPONT, N.; JAQUES, N.; WALDMAN-BROWN, A.; LUCCIONI, A. S.; MAHARAJ, T.; SHERWIN, E. D.; MUKKAVILLI, S. K.; KORDING, K. P.; GOMES, C. P.; NG, A. Y.; HASSABIS, D.; PLATT, J. C.; CREUTZIG, F.; CHAYES, J.; BENGIO, Y. Tackling climate change with machine learning. **ACM Comput. Surv.**, Association for Computing Machinery, New York, NY, USA, v. 55, n. 2, fev. 2022. ISSN 0360-0300. Disponível em: <https://doi.org/10.1145/3485128>.

RUMELHART, D. E.; HINTON, G. E.; WILLIAMS, R. J. Learning representations by back-propagating errors. **Nature**, v. 323, n. 6088, p. 533–536, Oct 1986. ISSN 1476-4687. Disponível em: <https://doi.org/10.1038/323533a0>.

SEMENZA, J. C. Cascading risks of waterborne diseases from climate change. **Nat Immunol**, United States, v. 21, n. 5, p. 484–487, maio 2020.

SEMENZA, J. C.; KO, A. I. Waterborne diseases that are sensitive to climate variability and climate change. **New England Journal of Medicine**, v. 389, n. 23, p. 2175–2187, 2023. Disponível em: <https://www.nejm.org/doi/full/10.1056/NEJMra2300794>.

SHAO, C.; ZHANG, H. Climate change characteristics and population health impact factors using deep neural network and hyperautomation mechanism. **The Journal of Supercomputing**, v. 80, n. 7, p. 8637–8667, May 2024. ISSN 1573-0484. Disponível em: <https://doi.org/10.1007/s11227-023-05795-y>.

SHARMA, S.; KUMAR, R. Pollen allergy prediction: A review of the time-frame-specific forecast models. In: **2023 Global Conference on Information Technologies and Communications (GCITC)**. [S.l.: s.n.], 2023. p. 1–6.

SINGH, P.; SARAN, S. Identifying and predicting climate change impact on vector-borne disease using machine learning: Case study of plasmodium falciparum from africa. **The International Archives of the Photogrammetry, Remote Sensing and Spatial Information Sciences**, XLVIII-2-2024, p. 387–391, 2024. Disponível em: <https://isprs-archives.copernicus.org/articles/XLVIII-2-2024/387/2024/>.

SJOEBERG, D.; HANNAY, J.; HANSEN, O.; KAMPENES, V.; KARAHASANOVIC, A.; LIBORG, N.-K.; REKDAL, A. A survey of controlled experiments in software engineering. **IEEE Transactions on Software Engineering**, v. 31, n. 9, p. 733–753, 2005.

SKIERA, B.; REINER, J.; ALBERS, S. Regression analysis. In: _____. **Handbook of Market Research**. Cham: Springer International Publishing, 2022. p. 299–327. ISBN 978-3-319-57413-4. Disponível em: <https://doi.org/10.1007/978-3-319-57413-4_17>.

STEFANOVIC, D.; HAVZI, S.; NIKOLIC, D.; DAKIC, D.; LOLIC, T. Analysis of the tools to support systematic literature review in software engineering. **IOP Conference Series: Materials Science and Engineering**, IOP Publishing, v. 1163, n. 1, p. 012013, aug 2021. Disponível em: <https://dx.doi.org/10.1088/1757-899X/1163/1/012013>.

TALLAM, K.; QUANG, M. P. Applications of artificial intelligence in predicting dengue outbreaks in the face of climate change: a case study along coastal india. **medRxiv**, Cold Spring Harbor Laboratory Press, 2023. Disponível em: <https://www.medrxiv.org/content/early/2023/01/23/2023.01.18.23284134>.

TIAN, N.; ZHENG, J.-X.; LI, L.-H.; XUE, J.-B.; XIA, S.; LV, S.; ZHOU, X.-N. Precision prediction for dengue fever in singapore: A machine learning approach incorporating meteorological data. **Tropical Medicine and Infectious Disease**, v. 9, n. 4, 2024. ISSN 2414-6366. Disponível em: <https://www.mdpi.com/2414-6366/9/4/72>.

TIKKANEN, R.; OSBORN, R.; MOSSIALOS, E.; DJORDJEVIC, A.; WHARTON, G. **International Profiles of Health Care Systems 2020**. [S.l.: s.n.], 2020.

VANHALA, L.; CALLIARI, E.; THOMAS, A. Understanding the Politics and Governance of Climate Change Loss and Damage. **Global Environmental Politics**, v. 23, n. 3, p. 1–11, 08 2023. ISSN 1526-3800. Disponível em: <https://doi.org/10.1162/glep_e_00735>.

VUURST, P. Van de; ESCOBAR, L. E. Climate change and infectious disease: a review of evidence and research trends. **Infectious Diseases of Poverty**, v. 12, n. 1, p. 51, May 2023. ISSN 2049-9957. Disponível em: <https://doi.org/10.1186/s40249-023-01102-2>.

WANG, J.; ZHANG, Y.; XU, Y. Editorial: Machine learning in studies of atmospheric environment and climate change. **Frontiers in Environmental Science**, v. 11, 2023.

WANG, X.; MA, Z.; DONG, J. Quantitative impact analysis of climate change on residents' health conditions with improving eco-efficiency in china: A machine learning perspective. **International Journal of Environmental Research and Public Health**, v. 18, n. 23, 2021. ISSN 1660-4601. Disponível em: <https://www.mdpi.com/1660-4601/18/23/12842>.

WATTS, N.; AMANN, M.; ARNELL, N.; AYEB-KARLSSON, S.; BELESOVA, K.; BERRY, H.; BOULEY, T.; BOYKOFF, M.; BYASS, P.; CAI, W.; CAMPBELL-LENDRUM, D.; CHAMBERS, J.; DALY, M.; DASANDI, N.; DAVIES, M.; DEPOUX, A.; DOMINGUEZ-SALAS, P.; DRUMMOND, P.; EBI, K. L.; EKINS, P.; MONTOYA, L. F.; FISCHER, H.; GEORGESON, L.; GRACE, D.; GRAHAM, H.; HAMILTON, I.; HARTINGER, S.; HESS, J.; KELMAN, I.; KIESEWETTER, G.; KJELLSTROM, T.; KNIVETON, D.; LEMKE, B.; LIANG, L.; LOTT, M.; LOWE, R.; SEWE, M. O.; MARTINEZ-URTAZA, J.; MASLIN, M.; MCALLISTER, L.; MIKHAYLOV, S. J.; MILNER, J.; MORADI-LAKEH, M.; MORRISSEY, K.; MURRAY, K.; NILSSON, M.; NEVILLE, T.; ORESZCZYN, T.; OWFI, F.; PEARMAN, O.; PENCHEON, D.; PYE, S.; RABBANIHA, M.; ROBINSON, E.; ROCKLÖV, J.; SAXER, O.; SCHÜTTE, S.; SEMENZA, J. C.; SHUMAKE-GUILLEMOT, J.; STEINBACH, R.; TABATABAEI, M.; TOMEI, J.; TRINANES, J.; WHEELER, N.; WILKINSON, P.; GONG, P.; MONTGOMERY, H.; COSTELLO, A. The 2018 report of the lancet countdown on health and climate change: shaping the health of nations for centuries to come. **The Lancet**, Elsevier, v. 392, n. 10163, p. 2479–2514, Dec 2018. ISSN 0140-6736. Disponível em: <https://doi.org/10.1016/S0140-6736(18)32594-7>.

WHEELER, R. L.; MILLER, J. S.; SMITH, G. N.; LAMBERT, J. H.; PENNETTI, C. A.; LOUIS, G. E.; JOHNSON, D. A. Geospatial analysis of extreme temperature impacts in agricultural systems using machine learning. In: **2024 IEEE International Symposium on Systems Engineering (ISSE)**. [S.l.: s.n.], 2024. p. 1–8.

WHITEFORD, L. M.; VINDROLA-PADROS, C. Comparative health systems. In: WRIGHT, J. D. (Ed.). **International Encyclopedia of the Social Behavioral Sciences (Second Edition)**. Second edition. Oxford: Elsevier, 2015. p. 400–404. ISBN 978-0-08-097087-5. Disponível em: <https://www.sciencedirect.com/science/article/pii/B9780080970868140097>.

WILLOX, A. C.; HARPER, S. L.; FORD, J. D.; EDGE, V. L.; LANDMAN, K.; HOULE, K.; BLAKE, S.; WOLFREY, C. Climate change and mental health: an exploratory case study from rigolet, nunatsiavut, canada. **Climatic Change**, v. 121, n. 2, p. 255–270, Nov 2013. ISSN 1573-1480. Disponível em: <https://doi.org/10.1007/s10584-013-0875-4>.

WILSON, J. R.; LORENZ, K. A. Introduction to binary logistic regression. In: _____. **Modeling Binary Correlated Responses using SAS, SPSS and R**. Cham: Springer International Publishing, 2015. p. 3–16. ISBN 978-3-319-23805-0. Disponível em: <https://doi.org/10.1007/978-3-319-23805-0_1>.

XU, N.; ZHANG, Y.; DU, C.; SONG, J.; HUANG, J.; GONG, Y.; JIANG, H.; TONG, Y.; YIN, J.; WANG, J.; JIANG, F.; CHEN, Y.; JIANG, Q.; DONG, Y.; ZHOU, Y. Prediction of oncomelania hupensis distribution in association with climate change using machine learning models. **Parasites & Vectors**, v. 16, n. 1, p. 377, Oct 2023. ISSN 1756-3305. Disponível em: <https://doi.org/10.1186/s13071-023-05952-5>.

ZAFEIROPOULOS, C.; TEMENOS, A.; ANGELI, A.; GIAKOUMAKI, A.; GALLOS, P.; KATSORIDA, E.; DOULAMIS, A. Toward health-centered monitoring framework promoting well-being and nature-based solutions. In: **Proceedings of the 17th International Conference on PErvasive Technologies Related to Assistive Environments**. New York, NY, USA: Association for Computing Machinery, 2024. (PETRA '24), p. 435–439. ISBN 9798400717604. Disponível em: <https://doi.org/10.1145/3652037.3663929>.

ZHANG, L.; JING, Z.; LI, Z.; LIU, Y.; FANG, S. Predictive modeling of suitable habitats for cinnamomum camphora (l.) presl using maxent model under climate change in china. **International Journal of Environmental Research and Public Health**, v. 16, n. 17, 2019. ISSN 1660-4601. Disponível em: <https://www.mdpi.com/1660-4601/16/17/3185>.