

MINERAÇÃO DE DADOS E PREVISÃO DA ECOTOXICIDADE DE FLUIDOS DE PERFURAÇÃO – UMA REVISÃO SISTEMÁTICA

DATA MINING AND ECOTOXICITY PREDICTION OF DRILLING FLUIDS – A SYSTEMATIC REVIEW

MINERÍA DE DATOS Y PREDICCIÓN DE LA ECOTOXICIDAD DE LOS FLUIDOS DE PERFORACIÓN - UNA REVISIÓN SISTEMÁTICA

LEOMIR SAMUEL TORMEN REIS | IFF - Instituto Federal Fluminense, Brasil

VICTOR BARBOSA SARAIVA, Dr. | IFF - Instituto Federal Fluminense, Brasil

SIMONE VASCONCELOS SILVA, Dra. | IFF - Instituto Federal Fluminense, Brasil

RESUMO

O comportamento ecotoxicológico de fluidos de perfuração pode ser estudado pela relação entre a ecotoxicidade de um fluido com suas propriedades físico-químicas e a concentração de seus componentes. Por se tratar de uma mistura complexa, os componentes do fluido influenciam de formas diferentes na ecotoxicidade da mistura. De posse de uma base de dados contendo resultados de testes de ecotoxicidade de amostras de fluidos de perfuração conhecidas, pode-se aplicar técnicas computacionais, como a Mineração de Dados, objetivando a previsão e extração de conhecimento, que podem ser empregadas em ferramentas de análises de risco e tomada de decisão. Este artigo tem como objetivo elaborar uma revisão sistemática da literatura para responder à questão de pesquisa: Quais são as técnicas computacionais mais utilizadas para a previsão e/ou conhecimento do comportamento toxicológico de fluidos de perfuração, conhecida a sua composição e suas propriedades físico-químicas? Para tanto, foram utilizadas três bases de dados e os resultados mostraram que a toxicologia computacional tem sido amplamente utilizada na previsão de toxicidade, contudo não há estudos relacionados a fluidos de perfuração nas bases de dados utilizadas. Foi possível concluir que para a previsão e/ou conhecimento do comportamento toxicológico dos fluidos são aplicadas técnicas computacionais de Mineração de Dados, utilizando a tarefa de Classificação por meio dos algoritmos Support Vector Machines e Random Forest.

PALAVRAS-CHAVE

Mineração de dados; ecotoxicidade; toxicologia; toxicidade; fluidos de perfuração.

ABSTRACT

The ecotoxicological behavior of drilling fluids can be studied by the relationship between the fluido ecotoxicity with its physical-chemical properties and the components's concentrations. Because it is a complex mixture, the components of the fluid influence the ecotoxicity of the mixture in different ways. With a database containing ecotoxicity test results of known drilling fluid samples, computational techniques such as Data Mining can be applied, aiming at forecasting and extracting knowledge, which can be used in analysis tools risk and decision making. This article aims to elaborate a systematic review of the literature to answer the research question: What are the most used computational techniques



for the prediction and/or knowledge of the toxicological behavior of drilling fluids, knowing their composition and their physical-chemical properties? For that, three databases were used and the results showed that computational toxicology has been widely used in the prediction of toxicity, however there are no studies related to drilling fluids in the databases used. It was possible to conclude that for the prediction and/or knowledge of the toxicological behavior of the fluids, computational techniques of Data Mining are applied, using the Classification task through the Support Vector Machines and Random Forest algorithms.

KEYWORDS

Data mining; ecotoxicity; toxicology; toxicity; drilling fluids.

RESUMEN

El comportamiento ecotoxicológico de los fluidos de perforación puede estudiarse mediante la relación entre la ecotoxicidad de un fluido con sus propiedades fisicoquímicas y la concentración de sus componentes. Al tratarse de una mezcla compleja, los componentes del fluido influyen de distintas maneras en la ecotoxicidad de la mezcla. En posesión de una base de datos que contenga los resultados de las pruebas de ecotoxicidad de muestras conocidas de fluidos de perforación, se pueden aplicar técnicas de minería destinadas a la predicción y la extracción de conocimientos, que pueden emplearse en herramientas de análisis de riesgos y toma de decisiones. Este trabajo pretende elaborar una revisión sistemática de la literatura para responder a la pregunta de investigación: ¿cuál es la predicción del comportamiento toxicológico de los fluidos de perforación, conocida su composición y sus propiedades fisicoquímicas, mediante técnicas computacionales de análisis de datos de ensayos de ecotoxicidad aguda? Para ello, se utilizaron tres bases de datos y los resultados mostraron que la toxicología computacional (in silico) se ha utilizado ampliamente en la predicción de la toxicidad, sin embargo no hay estudios relacionados con los fluidos de perforación en las bases de datos utilizadas.

PALABRAS CLAVE

Minería de datos, ecotoxicidad; toxicología; toxicidad; fluidos de perforación.



1. INTRODUÇÃO

Fluidos de perfuração são utilizados na perfuração de poços de petróleo, terrestres ou marítimos, e possuem diversas funções tais como: transportar os fragmentos de rocha perfurados (cascalhos) até a superfície da plataforma de perfuração; promover a estabilidade do poço; controlar a pressão em subsuperfície e conferir lubrificidade às ferramentas de perfuração (BAKER HUEGUES, 2016).

Estes fluidos são, em geral, constituídos por uma mistura de produtos químicos que tem por finalidade conferir as propriedades físico-químicas necessárias aos fluidos para que atendam às funções requeridas na perfuração. No que diz respeito à composição, os fluidos de perfuração podem ser de base aquosa ou não aquosa. Além disto, a formulação (composição) destes fluidos deve estar de acordo com diversos critérios ambientais para seu uso e descarte, assim como os cascalhos associados aos fluidos. Estes critérios são, no Brasil, normatizados pelo Instituto Brasileiro do Meio Ambiente e dos Recursos Naturais Renováveis (IBAMA) (BRASIL, 2018; BRASIL, 2019).

Os critérios para uso e descarte de fluidos de perfuração atualmente adotados para a perfuração de poços marítimos no Brasil estão apresentados pelo IBAMA no Despacho nº 5540547/2019 (BRASIL, 2019). Um destes critérios é a ecotoxicidade aguda, que deve ser determinada por meio de teste laboratorial seguindo os procedimentos apresentados normas NBR 15308 (ABNT, 2017) e NBR 15469 (ABNT, 2015) e está relacionada diretamente à composição dos fluidos, uma vez que os seus componentes conferem maior ou menor grau de toxidez à mistura.

Conforme esta metodologia, o teste consiste na exposição de uma série de diluições do material a ser analisado à um organismo teste. Neste caso específico, podem ser utilizados organismos *Mysidopsis juniae* e *Mysidium gracile* (ABNT, 2017). A Figura 01 apresenta um desenho esquemático da realização do teste de ecotoxicidade aguda.

Ao final do teste é calculada a porcentagem de letalidade para cada uma das diluições e, assim, determinada a concentração letal (CL) da amostra na qual ocorreu a letalidade de 50% dos organismos testes, após 96 horas de exposição, chamada de CL50-96h.

O valor da CL50-96h é o critério adotado pelo IBAMA no Despacho nº 5540547/2019 e deve ser igual ou superior a 30.000 ppm (30.000 mg/L). Neste caso, as amostras não são consideradas tóxicas e, conseqüentemente, obedecem ao critério de ecotoxicidade para uso e descarte de fluidos de perfuração e cascalhos associados a estes fluidos. Caso seja menor que 30.000 ppm, a amostra é

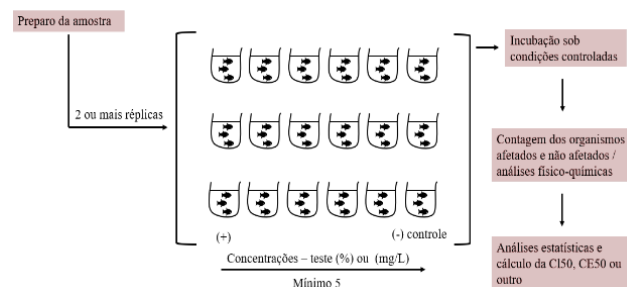


Figura 01: Desenho esquemático de um ensaio de toxicidade.

Fonte: VEIGA (2010).

considerada tóxica e neste caso, o fluido e os cascalhos associados não podem ser descartados no ambiente marinho (BRASIL, 2019).

Cabe ressaltar que a ecotoxicidade é apenas um dos critérios do monitoramento ambiental da atividade e, portanto, não pode ser utilizado isoladamente para determinação da permissão do uso e descarte de fluidos de perfuração.

Portanto, para atender aos requisitos ambientais e ao monitoramento ambiental da perfuração de poços, amostras dos fluidos de perfuração utilizados são encaminhadas aos laboratórios para a realização do teste de ecotoxicidade aguda e assim é determinada a CL50-96h de cada um dos fluidos utilizados. Contudo, os testes são realizados após a utilização e descarte dos fluidos e cascalhos e, em caso de reprovação, ou seja, constatada a CL50-96h menor que 30.000 ppm, o fluido não pode ser descartado no mar.

Logo, considerando-se que a ecotoxicidade aguda dos fluidos de perfuração está associada aos componentes deste fluido e de suas propriedades físico-químicas, é relevante o estudo da influência de cada um destes componentes, ou da mistura como um todo, na ecotoxicidade aguda apresentada pelo fluido de perfuração.

Em uma revisão sistemática, Aslan (2018) mostrou que muitos estudos avaliam a influência de determinados componentes do fluido em sua toxicidade. Outros, como o apresentado por Veiga (1998), avaliam os efeitos ecotoxicológicos de um mesmo fluido em diferentes organismos. Entretanto, poucos estudos avaliam a resposta ecotoxicológica considerando um maior número de variáveis, o que é relevante tendo em vista que os fluidos de perfuração são constituídos por uma mistura de produtos químicos.

Assim, de posse de uma base de dados contendo os componentes utilizados na formulação dos fluidos de perfuração, suas concentrações e o resultado dos testes de ecotoxicidade já realizados, pretende-se utilizar a mineração de dados para averiguar a relação entre as

concentrações dos componentes do fluido, assim como suas propriedades físico-químicas e o CL50-96h, buscando gerar conhecimento e a possibilidade de prever o valor do CL50-96h, conhecendo-se a composição do fluido, sem a necessidade de realização do teste de laboratório.

Esta previsão pode ser realizada em momento anterior à utilização do fluido, evitando ou diminuindo as chances de utilização de algum fluido de perfuração que venha a ser considerado tóxico após a realização do teste laboratorial de ecotoxicidade aguda.

A previsão do comportamento toxicológico de substâncias químicas por métodos computacionais de análises de dados vem sendo amplamente estudada recentemente. Esta área de estudo é conhecida como toxicologia computacional e lida com três desafios principais: i) modelagem de processos contínuos de substâncias químicas de fontes de que possuam efeito ecológico adverso; ii) avaliação de riscos de utilização de diferentes substâncias e iii) previsão do risco de substâncias químicas expostas a organismos de diversas espécies (TANG et al., 2019).

Na toxicologia computacional é possível fornecer ferramentas úteis para tomada de decisão por meio da utilização da modelagem computacional e de técnicas de mineração de dados, podendo reduzir ou eliminar a necessidade de realização dos testes tradicionais, como toxicidade aguda, diminuindo, portanto, a necessidade de testes em animais (MISHRA; FEI; HUAN, 2013).

A mineração de dados (MD) pode ser descrita como uma exploração de base de dados por meio do uso de tarefas adequadas para extração de conhecimento no formato de padrões e regras (FERRARI e SILVA, 2017). Nesse sentido, Lorenzetti e Teloken (2016) destacam que a MD é composta por diversas tarefas criadas para facilitar a descoberta de algo novo e que possa contribuir no conhecimento a partir de milhares de informações. Desse modo, os algoritmos de MD são baseados em um tipo de tarefa.

As tarefas mais utilizadas são classificação, regressão, clusterização e associação (WITTEN; FRANK; HALL, 2016): (i) classificação é o processo de encontrar um modelo que descreve e distingue classes ou conceitos de dados; (ii) regressão modela funções de valor contínuo, envolvendo a identificação de tendências de distribuição com base nos dados disponíveis; (iii) clusterização é descrito como um conjunto de dados, onde é possível agrupar os dados iguais juntos em um mesmo grupo e os desiguais em grupos distintos; e (iv) associação tem como objetivo elaborar uma representação explícita entre os objetos, visando determinar relacionamentos entre conjuntos de itens de associação.

Nesse contexto, o objetivo inicial deste trabalho visava

a elaboração de uma revisão sistemática da literatura capaz de responder à questão de pesquisa geral: QP - Quais são as técnicas computacionais mais utilizadas para a previsão e/ou conhecimento do comportamento toxicológico de fluidos de perfuração, conhecida a sua composição e suas propriedades físico-químicas?

Para tal, foi realizada a busca na literatura nas bases de dados Scopus, Web of Science e Catálogo de Teses e Dissertações da CAPES. Esta busca foi realizada da mesma forma nas três bases, utilizando-se os seguintes tesouros: toxicidade, ecotoxicidade, fluidos de perfuração, mineração de dados e aprendizado de máquina.

O termo de busca, em todas as bases foi: (toxicity OR ecotoxicity) AND (“drilling fluid*” OR “drilling mud*”) AND (“data mining” OR “machine learning”). Este termo foi escolhido para ampliar os resultados aderentes ao tema em questão, uma vez que: i) o termo ecotoxicidade é também expresso na literatura como toxicidade ambiental ou apenas toxicidade; ii) o termo fluido de perfuração é também relatado como lama de perfuração em alguns casos e iii) mineração de dados e aprendizado de máquina são temas entrelaçados e uma análise minuciosa nos achados literários deve ser realizada para se determinar a aderência ao tema em questão.

A busca foi realizada no primeiro trimestre de 2022, e não retornou resultados, o que indica que não foi encontrado até o momento da busca estudos relacionando os três temas relevantes a esta pesquisa (ecotoxicidade - ou toxicidade, mineração de dados e fluidos de perfuração). Além disto, a ausência de resultados indica também uma lacuna na documentação científica relacionada à previsão da ecotoxicidade de fluidos de perfuração.

Diante deste cenário, o objetivo do presente trabalho foi alterado, provendo a elaboração de uma revisão sistemática da literatura a fim de responder as seguintes questões específicas de pesquisa: QEP1 – Como é feita a Ecotoxicologia de fluidos de perfuração? e QEP2 – Como a Mineração de Dados é utilizada para previsão da toxicidade? Por meio das respostas obtidas para as duas questões acima, este trabalho visa propor uma resposta para a questão de pesquisa geral proposta (QP).

2. MATERIAIS E MÉTODO

Foram realizadas buscas nas bases Scopus, Web of Science e Catálogo de Teses e Dissertações da CAPES. As buscas realizadas foram divididas em duas categorias, de modo a responder as duas questões de pesquisa propostas: i) para

QP1, utilizando os seguintes termos: (toxicity OR ecotoxicity) AND ("drilling fluid*" OR "drilling mud*"); e ii) para QP2, utilizando os seguintes termos: ("toxicity prediction" OR "ecotoxicity prediction") AND ("data mining" OR "machine learning").

Optou-se pela busca do termo já combinado “previsão de toxicidade (ou ecotoxicidade)” em função de que estudos relacionados à toxicidade de substâncias químicas voltadas à área médica estão amplamente descritos na literatura, sobretudo com foco em fármacos e, portanto, o termo escolhido foi utilizado para a obtenção de resultados que tratam da previsão de toxicidade.

A Figura 02 mostra um desenho esquemático do comportamento observado dos relatos encontrados na busca na literatura, no que diz respeito às interseções das áreas afins.

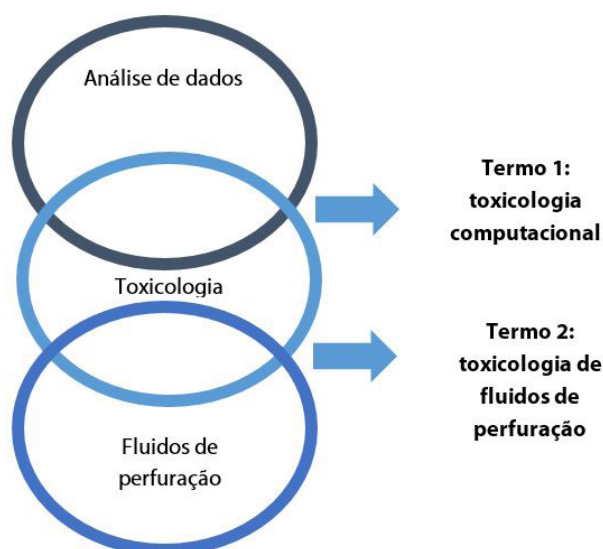


Figura 02: Análise dos resultados da busca na literatura.
Fonte: Autores.

Observando-se a Figura 02, percebe-se que estão documentadas na literatura as previsões ecotoxicológicas por meio de métodos computacionais de mineração de dados e aprendizado de máquina, bem como análises de estudos de dados ecotoxicológicos de fluidos de perfuração.

A interseção dos temas mineração de dados, aprendizado de máquina e ecotoxicologia, corresponde à área de previsão de toxicidade (ou toxicologia computacional), por meio da qual é possível a utilização de métodos computacionais para a previsão da ecotoxicidade de uma determinada amostra, seja substância pura ou mistura.

A interseção dos temas ecotoxicologia e fluidos de perfuração corresponde à ecotoxicologia de fluidos de perfuração, amplamente documentada tendo em vista que

a ecotoxicidade é um dos critérios para uso e descarte de fluidos de perfuração no Brasil e no mundo.

2.1 Ecotoxicologia de fluidos de perfuração

A busca dos dados referentes à ecotoxicologia de fluidos de perfuração foi realizada com o seguinte termo de busca: (toxicity OR ecotoxicity) AND ("drilling fluid*" OR "drilling mud*"). Utilizou-se como critério para refinamento, a busca de relatos publicados nos últimos 20 anos e relatos que sejam artigos publicados em periódicos, teses ou dissertações, o que exclui resumos e artigos de conferências.

O filtro temporal de 20 anos foi utilizado em função do avanço das legislações ambientais a partir do ano 2000 e pelo fato de que pesquisas em anos anteriores podem estar defasadas em relação às legislações atuais, bem como em relação ao tema propriamente dito.

Utilizou-se a metodologia PRISMA (Preferred Reporting Items for Systematic Reviews and Meta-Analyses) (PAGE et al., 2021) adaptada para a seleção dos relatos relevantes e que foram lidos integralmente. A Figura 03 apresenta um desenho esquemático do método utilizado na busca dos dados da literatura, bem como os principais resultados e os refinamentos utilizados.

A análise dos títulos e resumos limitou-se a selecionar os relatos que apresentam a relação entre um ou mais componentes dos fluidos de perfuração com sua toxicidade e relatos que tratam de análise de risco e/ou tomada de decisão no que diz respeito à toxicidade de fluidos de perfuração.

Foram descartados estudos que tinham foco em fornecer formulações de fluidos de perfuração para determinado cenário operacional e que traziam pouca ou nenhuma informação toxicológica, mencionando apenas que a formulação proposta atende a determinados critérios toxicológicos.

Foram, então, selecionados 15 relatos para análise e discussão nesta revisão.

2.2 Previsão de toxicidade

A segunda busca por relatos na literatura foi realizada utilizando-se o seguinte termo de busca, em todas as bases: ("toxicity prediction" OR "ecotoxicity prediction") AND ("data mining" OR "machine learning"). Utilizou-se como critério para refinamento, a busca de relatos publicados nos últimos 10 anos e relatos que sejam artigos publicados em periódicos, excluindo-se, portanto, resumos e artigos

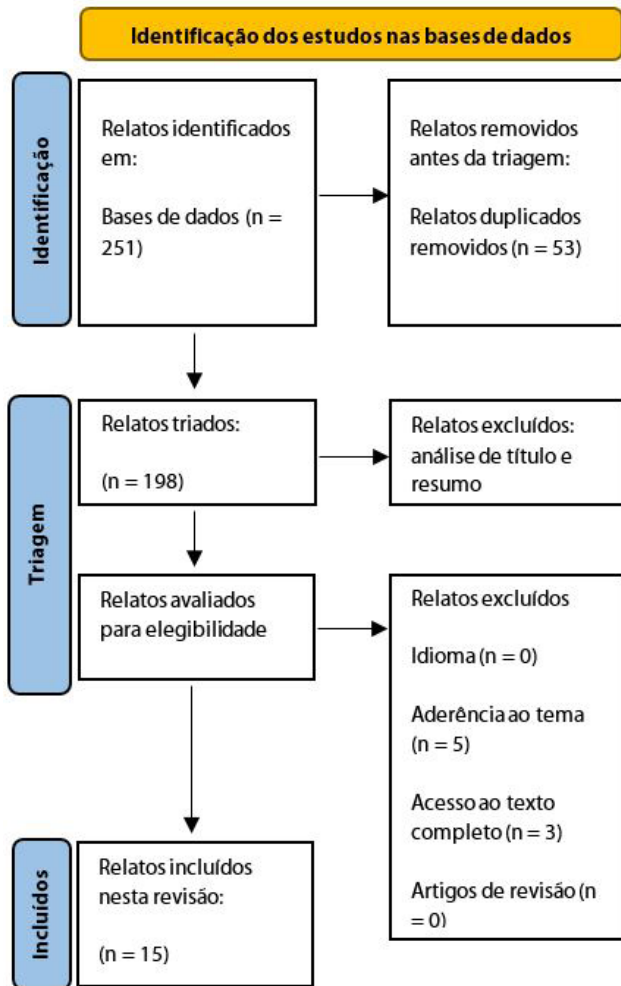


Figura 03: Método PRISMA adaptado (PAGE et. al, 2021). Ecotoxicologia de fluidos de perfuração.
Fonte: Autores.

de conferências.

O filtro temporal de 10 anos foi utilizado, pois, diferentemente da pesquisa relacionada à ecotoxicologia de fluidos de perfuração, a previsão de toxicidade por métodos computacionais é mais recente e está atrelada ao desenvolvimento e disponibilização destes métodos e por isto relatos mais antigos não foram considerados relevantes para este estudo.

Utilizou-se, também, a metodologia PRISMA (PAGE et. al., 2021) adaptada para a seleção dos relatos relevantes. A Figura 04 apresenta um

desenho esquemático do método utilizado. A análise dos títulos e resumos limitou-se a selecionar os relatos que apresentam estudos a respeito da previsão da toxicidade de substâncias químicas ou misturas frente a organismos testes, sendo, portanto, excluídos relatos da área da saúde, de toxicidade a humanos, notadamente relacionados a testes de novos fármacos.

Como pode ser observado na Figura 04, foram

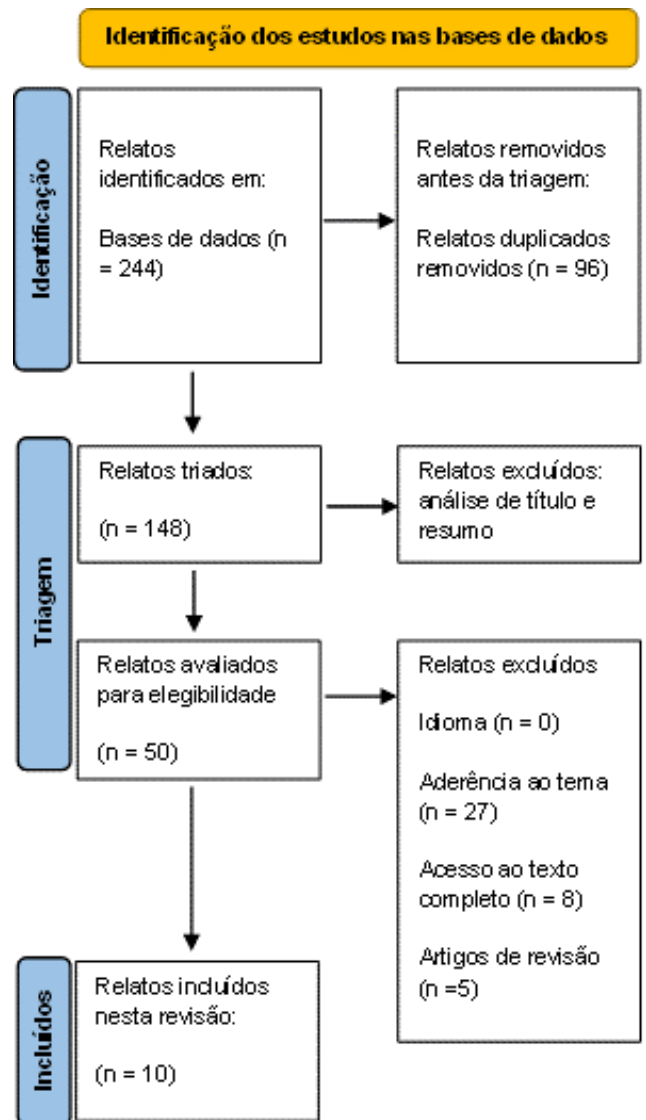


Figura 04: Método PRISMA adaptado (PAGE et. al, 2021). Previsão de toxicidade.
Fonte: Autores.

selecionados 10 relatos para análise e discussão. O número total de relatos obtidos da literatura e que contemplam estudos sobre a ecotoxicologia de fluidos de perfuração e sobre a previsão de toxicidade, considerando os critérios descritos no método de pesquisa realizado, e elegidos para análise e discussão nesta revisão é de 25 relatos.

3. RESULTADOS E DISCUSSÃO

Nesta seção serão apresentados os resultados dos relatos selecionados, assim como uma discussão dos principais resultados obtidos.

3.1 Ecotoxicologia de fluidos de perfuração

A Tabela 01 apresenta uma síntese dos principais resultados encontrados em cada estudo selecionado sobre

ecotoxicologia de fluidos de perfuração.

A Figura 05 mostra o número de relatos por ano de publicação. Observa-se um número crescente de publicações entre 2010 e 2012 e um número constante nos anos seguintes.

Um mapa esquemático do número de relatos por país de publicação é apresentado na Figura 06. Nota-se que os países que mais publicaram sobre a ecotoxicologia de fluidos de perfuração nos últimos anos foram a China (5 estudos), a Austrália (2 estudos) e a Nigéria (2 estudos).

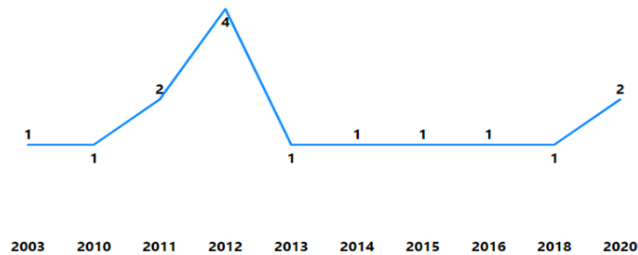


Figura 05: Número de relatos por ano de publicação.
 Fonte: Autores.

Aslan (2018) observou grande representatividade dos países Nigéria e Austrália no que diz respeito ao número de estudos publicados e atribuiu este fato a estes países estarem despontando como novas fronteiras exploratórias de petróleo e, portanto, precisam conhecer a ecotoxicologia dos fluidos de perfuração em relação aos organismos da fauna local para que sejam estabelecidos os critérios a serem adotados no monitoramento ambiental

a seguir.

A Figura 07 mostra a relação entre os dados extraídos dos relatos analisados, no que diz respeito à substância testada e ao tipo de teste ecotoxicológico realizado em cada um deles.

Observa-se que apenas 11,1% trataram de toxicidade crônica, sendo os demais relacionados a toxicidade aguda (44,4%) e toxicidade aguda e crônica (44,4%). Já em relação aos estudos que utilizaram o componente como substância testada, 75% avaliaram a toxicidade crônica,

enquanto os outros 25% trataram de toxicidade aguda

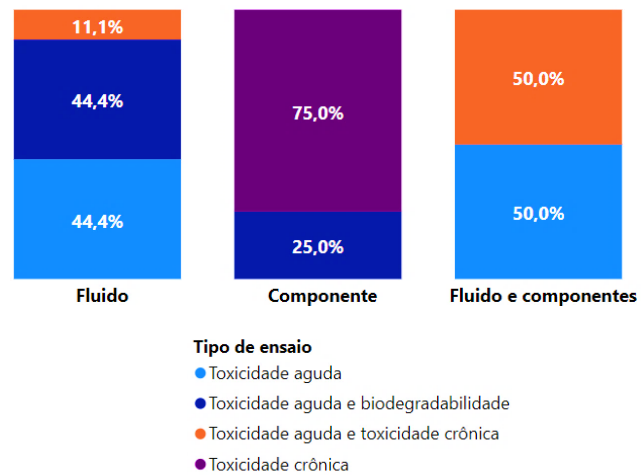


Figura 07: Relação entre os tipos de teste e a substância testada.
 Fonte: Autores.



Figura 06: Mapa esquemático do número de relatos por país.
 Fonte: Autores.

da atividade de perfuração de poços de petróleo e gás destas localidades.

As Figuras 07, 08 e 09 contribuem para responder à questão de pesquisa QE1 e são apresentadas e descritas

e biodegradabilidade. No que se refere aos estudos que analisaram tanto o fluido como algum de seus componentes, metade deles avaliou a toxicidade aguda e crônica e a outra metade apenas toxicidade aguda.

Em relação à amostra testada nos estudos analisados (componente o fluido de perfuração), pode-se observar

Autores	Ano	País	Periódicos	Quartil	Substância Testada	Tipo de ensaio
Qu et. al.	2020	China	<i>Chemistry and Technology of Fuels and Oils</i>	Q3	Componente	Toxicidade aguda e biodegradabilidade
Sun et. al.	2020	China	<i>Frontiers in Chemistry</i>	Q1	Fluido	Toxicidade aguda
Aslan	2018	Brasil	Catálogo de teses e dissertações	----	Fluido e componentes	Toxicidade aguda e toxicidade crônica
Chen et. al.	2016	China	<i>Journal of Cleaner Production</i>	Q1	Fluido e componentes	Toxicidade aguda
Durgut et. al.	2015	Noruega	<i>Marine Pollution Bulletin</i>	Q1	Fluido	Toxicidade aguda e toxicidade crônica
Steliga, Uliasz	2014	Polônia	<i>Mineral Resources Management</i>	Q3	Fluido	Toxicidade aguda e biodegradabilidade
Gagnon, Bakhtyar	2013	Austrália	<i>PLoS ONE</i>	Q1	Componente	Toxicidade crônica
Bakhtyar, Gagnon	2012	Austrália	<i>Environmental Monitoring and Assessment</i>	Q2	Componente	Toxicidade crônica
Sil et. al.	2012	Índia	<i>Journal of Hazardous, Toxic, and Radioactive Waste</i>	Q2	Fluido	Toxicidade aguda e biodegradabilidade
Ge et. al.	2012	China	<i>Petroleum Science and Technology</i>	Q2	Fluido	Toxicidade aguda e biodegradabilidade
Xie et. al.	2012	China	<i>Petroleum science and technology</i>	Q2	Fluido	Toxicidade aguda e biodegradabilidade
Lira et. al.	2011	Bélgica e Brasil	<i>Marine Environmental Research</i>	Q3	Componente	Toxicidade crônica
Ogeleka, Tudararo-Aherobo	2011	Nigéria	<i>African journal of aquatic science</i>	Q3	Fluido	Toxicidade aguda
Favour, Vincent-Akpu, et. al.	2010	Nigéria	Ciência Rural	Q3	Fluido	Toxicidade aguda
Sadiq et. al.	2003	Canadá	<i>International Journal of Environmental Studies</i>	Q3	Fluido	Toxicidade aguda

Tabela 01: Principais informações dos relatos analisados sobre ecotoxicologia de fluidos de perfuração.
Fonte: Autores.

na Figura 08 que 60% dos estudos tiveram foco na avaliação da ecotoxicidade do fluido e 26,7% na de seus componentes. Além destes, 13,3% dos estudos avaliaram a toxicidade tanto do fluido quanto de seus componentes.

Alguns estudos que tratam da análise de componen-

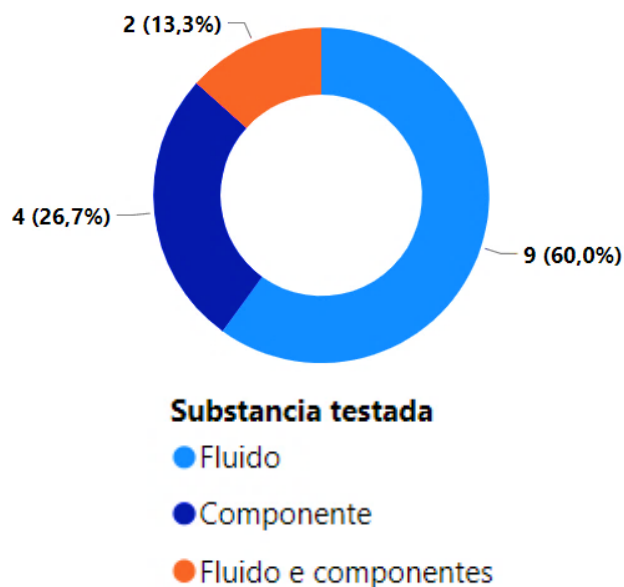


Figura 08: Número de relatos por tipo de substância.
Fonte: Autores

tes de fluido de perfuração têm como objetivo principal a avaliação da toxicidade deste componente em uma formulação de fluidos que atenda aos requisitos ambientais, ou seja, avaliar a performance de um fluido de perfuração que atenda aos critérios operacionais e seja ambientalmente correto.

O estudo de Qu et. al (2020) mostrou que a toxicidade de aminas orgânicas diminui de acordo com a conformação carbônica desta amina, ou seja, aminas primárias são menos tóxicas que aminas secundárias que por sua vez são menos tóxicas que aminas terciárias; além disto, mostraram também que a biodegradabilidade também diminui com o aumento do número de carbonos da conformação, sendo aminas primárias mais biodegradáveis que aminas secundárias e terciárias. Este mesmo estudo mostrou que as aminas são menos tóxicas e mais biodegradáveis que as amidas, ácidos e álcoois. Aminas e amidas são frequentemente utilizadas como agentes anti-colapso em fluidos de perfuração.

Gagnon e Bakhtyar (2013) mostraram que diferentes bases sintéticas apresentam toxicidade e biodegradabilidade diferentes. Assim, a base éster é mais biodegradável que as bases olefina interna e alfa olefina, por possuir

menor peso molecular, entretanto a base apresentou maior toxicidade dentre as três. Este estudo destaca também que a exposição real (em condições de campo) dos organismos da fauna local aos compostos testados não pode ser diretamente extrapolada pela análise dos dados apresentados, que foram obtidos em laboratório.

Sun et. al. (2020) sintetizaram uma base capaz de conferir boa performance operacional a um fluido de perfuração utilizando esta base, bem como aceitação aos critérios de ecotoxicidade. A base sintética em questão possui cadeias carbônicas constituídas de 12 a 22 carbonos e apresenta 10% da toxicidade do diesel.

Sadiq et. al. (2003) mostraram que a base sintética estudada por eles é menos tóxica do que a base óleo, já em desuso. Entretanto, além da toxicidade da base sintética deve-se considerar

também a presença de metais pesados como bário, presente na baritina, e de óleo da formação, presente no cascalho na avaliação do descarte de fluidos e cascalhos no ambiente marinho.

Aslan (2018), em sua revisão sistemática da literatura, mostrou que a baritina (sulfato de bário), um dos componentes dos fluidos de perfuração (adensante), contribui para o aumento da toxicidade destes fluidos.

Este comportamento foi relatado também por Chen et. al. (2016), Sadiq et. al (2003) e Lira et. al. (2011), sendo que neste último estudo foi mostrado que concentrações de bário inferiores a 300 ppm não conferiram efeitos toxicológicos à espécie testada (*Rhabditis (Pellioditis) marina*), entretanto, concentrações entre 400 e 2000 ppm conferiram este efeito, sendo observada mortalidade total dos organismos testes na concentração de 4800 ppm de bário.

A alcalinidade do fluido também influencia em sua toxicidade (ASLAN, 2018). Fluidos de perfuração, em sua maioria, possuem pH alcalino tendo em vista a degradação do próprio fluido em meio ácido. (GE et. al, 2012). Aslan (2018) realizou uma série de testes para avaliar a influência da alcalinidade na toxicidade de fluidos de perfuração e concluiu que a toxicidade aumenta com o aumento do pH e os maiores valores de CL50-96h foram obtidos com pH em torno da neutralidade.

Um dos componentes dos fluidos de perfuração aquosos é o bactericida, que possui função controlar o crescimento de bactérias no fluido, utilizando os polímeros ou outros componentes do fluido como nutrientes, o que pode degradá-lo.

Steliga e Uliasz (2014) mostraram que a concentração de bactericida no fluido de perfuração influencia em sua

biodegradabilidade, de forma que quanto maior for a concentração de bactericida, menor é a biodegradabilidade do fluido.

Bakhtyar e Gagnon (2012) evidenciaram que dos muitos componentes do fluido de perfuração, alguns não influenciam na toxicidade, tais como o agente de controle de filtrado e o viscosificante utilizados em seu estudo. Entretanto, outros componentes estão diretamente relacionados à toxicidade do fluido, tais como os emulsificantes, que causaram alterações na atividade da proteína EROD e dano ao DNA dos peixes (organismos testes) da espécie *Pagrus auratus*.

Também utilizando peixes como organismos teste, Vincent-Akpui, Allison e Sikoki (2010), utilizando a espécie de peixe *Tilapia guineenses*, estudaram o comportamento toxicológico de um fluido de perfuração em diferentes fases da vida destes peixes e concluíram que alevinos são mais sensíveis que larvas e juvenis à exposição ao mesmo fluido. Assim, evidenciou-se que a toxicidade dos fluidos pode desequilibrar a distribuição do número de organismos (peixes) de uma mesma espécie em diferentes estágios de vida. Isto fica ainda mais crítico quando há grande diversidade de espécies vivendo no mesmo local.

No mesmo sentido, o estudo realizado por Sil et. al. (2012), mostrou que diferentes espécies de peixes (*Tilapia mossambica* e *Mugil persia*) são sensíveis ao fluido de perfuração testado e que este fato demonstra a necessidade de uma diretriz ambiental para regulamentação da atividade de perfuração de poços marinhos na Índia.

Durgut et. al. (2015) concluíram em seus estudos realizados com organismos que vivem na camada de sedimentos (comunidade bentônica) que o teor de oxigênio dissolvido no sedimento diminui quando ocorre a disposição de cascalhos de perfuração no fundo do mar em cerca de 6 cm em relação ao sedimento original. Esta diminuição é consequência da degradação da matéria orgânica depositada juntamente com os cascalhos (fluido aderido) e que a concentração original de oxigênio dissolvido é retomada cerca de 1 ano após a deposição do resíduo.

Assim como dito no estudo de Sadiq et. al (2003) deve-se resguardar a comunidade bentônica tendo em vista que a toxicidade de

fluidos frente a estes organismos está diretamente relacionada às concentrações dos componentes dos fluidos de perfuração e, portanto, processos de separação sólido-líquido com foco na extração do fluido aderido ao cascalho podem contribuir na diminuição da concentração de fluido aderido ao cascalho descartado no mar.

A mesma observação está descrita por Ogeleka e Tudararo-Aherobo (2011) que enfatizam que organismos invertebrados que vivem na comunidade bentônica são fonte de alimento de outras espécies.

Devido à relação direta entre os aditivos do fluido de perfuração e a toxicidade frente a fauna bentônica, Xie et. al. (2012) estudou a possibilidade de disposição dos cascalhos com fluido aderido em um deserto na China, concluindo que após a disposição em uma camada me-

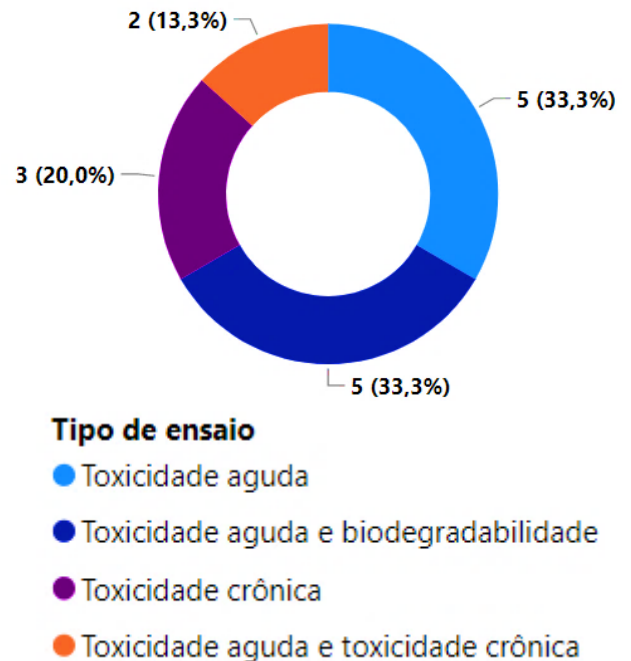


Figura 09: Número de relatos por tipo de ensaio ecotoxicológico.
Fonte: Autores

nor que 20 cm na superfície do solo, o fluido estudado aumentou a fertilidade do solo.

No que diz respeito ao tipo de ensaio toxicológico realizado nos estudos analisados, a Figura 09 mostra que 33,3% dos estudos analisaram a toxicidade aguda e 33,3% avaliaram tanto a toxicidade aguda quanto a biodegradabilidade. Além disto, 20% dos estudos consistiam na avaliação da toxicidade crônica e dois estudos (13,3%) avaliaram tanto a toxicidade aguda quanto a crônica.

A comparação entre toxicidade aguda e toxicidade crônica pode ser realizada nos estudos que contemplam os dois tipos de teste, utilizando os mesmos organismos testes.

Neste sentido, Aslan (2018) afirma que dos estudos analisados em seu trabalho, os fluidos não aquosos apresentaram baixa toxicidade aguda, porém se mostraram mais tóxicos quando se considera a toxicidade crônica, evidenciando que efeitos de longo prazo devem ser

considerados no monitoramento ambiental.

Durgut et al. (2015), com os estudos focados nos organismos bentônicos (que vivem no sedimento marinho), mostrou que o ensaio de toxicidade crônica, associado à biodegradabilidade, foi o melhor parâmetro para a modelagem global do efeito toxicológico sobre a fauna bentônica.

3.2 Previsão de Toxicidade

No que diz respeito à previsão de toxicidade, foram selecionados 10 relatos para leitura integral, por meio do método apresentado na Figura 04. As principais informações de cada um deles estão listadas na Tabela 02.

A Figura 10 relaciona o número de publicações por ano ao longo da última década. O número crescente de publicações pode estar associado ao ganho de relevância do assunto nos anos recentes, bem como à geração e armazenamento de grandes quantidades de dados que

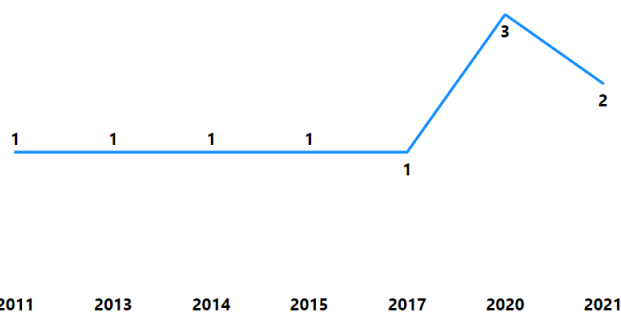


Figura 10: Número de relatos por ano.

Fonte: Autores.

necessitam de ferramentas estatísticas atreladas à métodos computacionais apropriados para serem analisados.

Este viés é corroborado por Tang et al. (2019), que afirmam que há dados ecotoxicológicos disponíveis em bases de dados públicas e que esta quantidade de dados é muito maior do que dados de testes tradicionais *in vivo* e *in vitro*, apresentando características de big data, tais como grandes volumes de dados e de informações complexas e interligadas. Além deste aspecto, diversos relatos citam o ganho de notoriedade e a necessidade do desenvolvimento de ferramentas para a previsão de toxicidade.

Pérez Santín et al. (2021) dizem que a modelagem do efeito nocivo de substâncias químicas ao ambiente para a previsão de toxicidade é de grande interesse, pois é crescente o debate relacionado a problemas de segurança e degradação ambiental com foco na necessidade de redução do impacto ambiental.

Mishra, Fei, Huan (2013), neste mesmo sentido, explicam que a EPA (Environmental Protection Agency) iniciou

em 2007 o projeto Tox21 (Toxicology in the twenty-first Century) que objetiva, entre outros, o desenvolvimento de assinaturas ou modelos preditivos (de toxicidade) que possam alcançar a mais alta acurácia quando comparados a qualquer método computacional existente, além de poderem ser aplicados para a previsão da toxicidade de substâncias químicas que ainda não foram exploradas, ou seja, não tiveram seu perfil toxicológico avaliado por métodos convencionais.

Como o recorte temporal utilizado para a busca de relatos científicos neste estudo foi de 10 anos, o projeto Tox21 foi lançado antes do período de análise desta revisão e, portanto, é citado em muitos dos relatos aqui analisados. Segundo Tang et al. (2019), o projeto Tox21 analisou cerca de 10000 substâncias químicas frente a mais de 70 tipos de avaliação toxicológica e, como resultado, disponibilizou estes dados que atualmente se encontram em diversas bases de dados públicas.

Surge então o termo Toxicologia Computacional (ou toxicologia *in silico*), que segundo a EPA é definido como uma combinação interdisciplinar de química ambiental, físico-química, toxicologia, bioinformática, quimioinformática, estatística, ciência da computação e outras áreas afins.

Yin, et al. (2011) utilizou ferramentas computacionais para construir um fluxo de trabalho eficiente para propostas de modelos de regressão quantitativos e confiáveis. Neste estudo, é afirmado que para de fato haver a substituição de testes de toxicidade tradicionais por metodologias *in silico*, estes devem possuir a mais alta eficácia e confiabilidade, além de não se limitar à previsão de e identificação do perigo de substâncias químicas (é ou não é tóxico), estabelecendo também a potencialidade deste perigo.

Em geral, a toxicologia computacional lida com três desafios: i) modelagem de processos contínuos de substâncias químicas de fontes que possuam efeito ecológico adverso; ii) avaliação de riscos de diferentes substâncias químicas e iii) previsão do risco de substâncias químicas quando expostas a diversas espécies (TANG et al., 2019).

Yin, et al. (2011) afirmam que para de fato haver a substituição de testes de toxicidade tradicionais por metodologias *in silico*, estes devem possuir a mais alta eficácia e confiabilidade, além de não se limitar à previsão de e identificação do perigo de substâncias químicas, estabelecendo também a potencialidade deste perigo.

Neste sentido, Ma, Buontempo, Wang (2011) mostraram que a mineração de dados é uma das técnicas de descoberta de conhecimento que podem ser

Autores	Ano	País	Periódico	Quartil	Base de Dados	Algoritmos / Modelos
Morger et al.	2021	Alemanha	<i>Journal of Cheminformatics</i>	Q1	TOXCast	Análise Conformacional
Fan et al.	2021	China	<i>Science of the Total Environment</i>	Q1	ECOTOX	SVM
Shadrin et al.	2020	Rússia	<i>Ecotoxicology and Environmental Safety</i>	Q1	Própria	SVM e ANN
Duan et al.	2020	China	<i>Talanta</i>	Q1	Própria	Regressão linear, kNN, SVM, GBRT e RF
Takata et al.	2020	Japão	<i>Chemosphere</i>	Q1	Diversas	Própria
Castillo-Garrit et al.	2017	Espanha	<i>SAR and QSAR in Environmental Research</i>	Q3	MOA Data set	kNN, SVM, ANN e DT.
Mishra, Fei, Huan	2013	Estados Unidos	<i>International Journal of Data Mining and Bioinformatics</i>	Q3	TOXCast	RF
Sherhod et al.	2012	Reino Unido	<i>Journal of Chemical Information and Modeling</i>	Q1	<i>Ames mutagenicity e hERG channel inhibition</i>	CP-tree
Ma, Buontempo, Wang	2011	Reino Unido	<i>Journal of Modelling, Identification and Control</i>	Q3	Própria	DT e DF
Yin, et al.	2011	China	<i>Molecular Informatics</i>	Q2	<i>Tetrahymena pyriformis toxicity dataset e Aqueous solubility dataset</i>	SVM

Legenda: kNN (k-Nearest Neighbors), SVM (Support Vector Machine), NB (Naïve Bayes), DT (Árvores de Decisão - Decision Trees), RF (Random Forest), GBRT (Gradient Boosted Regression Tress), CP-Tree (Compact Pattern Tree), NN (Redes Neurais), ANN (Redes Neurais Artificiais - Artificial Neural Network), DNN (Deep Neural Network).

Tabela 02: Principais dados dos estudos analisados, contendo as bases de dados e modelos ou algoritmos utilizados.

Fonte: Autores.

utilizadas neste contexto, sendo capazes de gerar bons modelos, fornecendo previsões quanti e qualitativas e

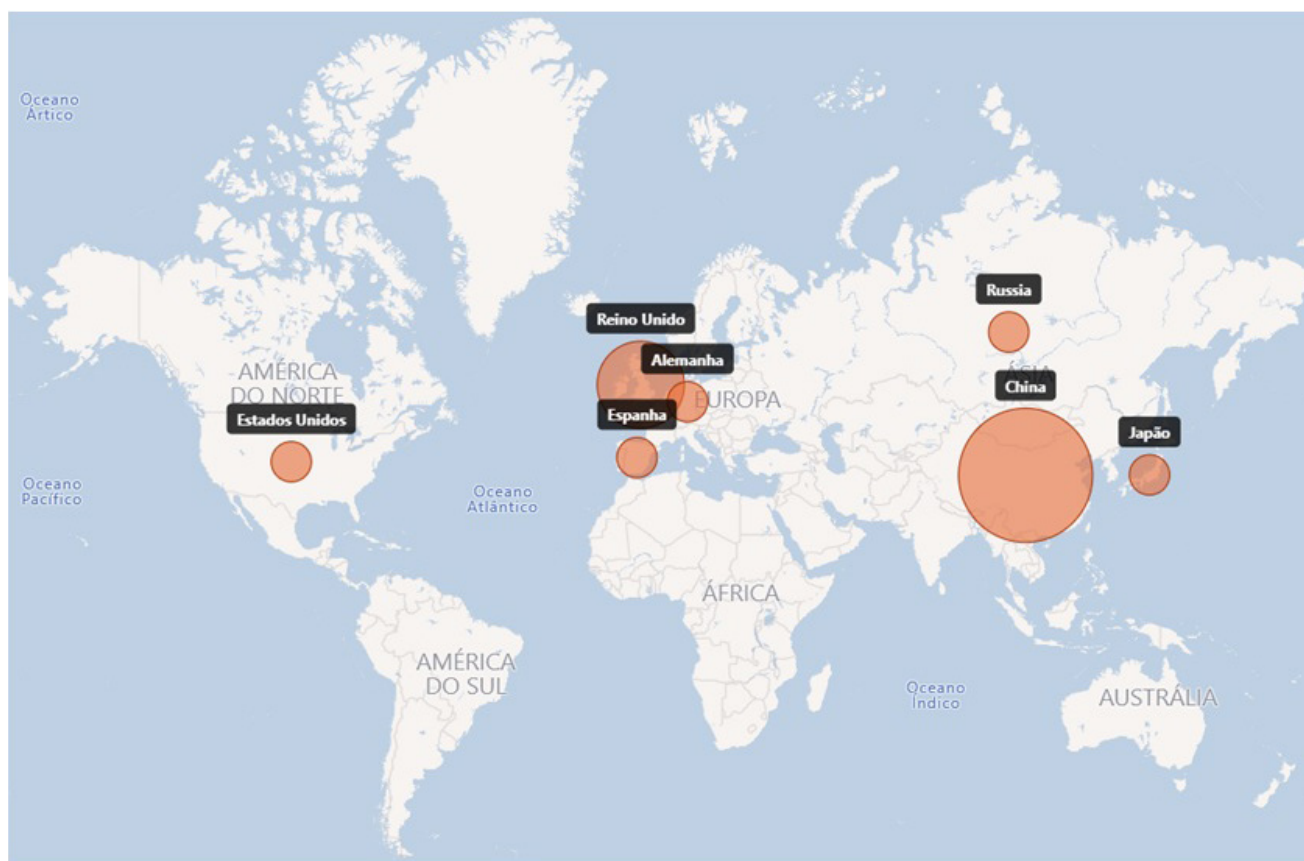


Figura 11: Mapa esquemático do número de relatos por país.

Fonte: Autores.

ainda incorporar conhecimentos previamente estabelecidos, o que fornece vantagens sobre modelos puramente numéricos.

Em relação aos países nos quais os estudos analisados foram desenvolvidos, há uma concentração em poucos países, apenas 7, como mostra a Figura 11, assim como pode-se também observar que esta mesma concentração de estudos realizados na Europa (5) e na China (3). Os países que complementam os demais estudos, todos estão localizados no hemisfério norte.

Uma explicação inicial para esta questão é o projeto To21 da EPA que alavancou os estudos relacionados à previsão de toxicidade no mundo todo. Observa-se, portanto, um predomínio de países desenvolvidos nas publicações aqui analisadas. Isto pode ser explicado pelo avanço legislativo de países desenvolvidos que tem passado a adotar metodologias computacionais de previsão e toxicidade para auxílio na tomada de decisão em detrimento aos métodos tradicionais de análise, seja por questões éticas, de custo ou disponibilidade de informações.

Neste sentido, Mishra, Fei, Huan (2013) citam em seu estudo que o processo de determinação da toxicidade tradicional, que corresponde a um processo no qual um grupo de organismos vivos é exposto a uma determinada

concentração de uma substância e a resposta destes organismos é medida após certo período, possui alto custo, seja este custo financeiro, de tempo ou ético (por envolver testes realizados com animais).

No debate relacionado à ética da realização de testes em animais, associado ao seu custo financeiro e de tempo, algumas autoridades regulatórias permitem a utilização de ferramentas computacionais para previsão e toxicidade e outras propriedades de substâncias químicas, quando aplicável, obedecendo alguns critérios.

É o caso da legislação REACH (Registration, Evaluation, Authorisation, Restriction of Chemicals), da União Europeia (região responsável pela publicação de 5 relatos aqui analisados). Isto se tornou mais factível após a publicação, pela OCDE (Organização para a Cooperação e Desenvolvimento Econômico), dos 5 princípios para utilização de técnicas computacionais em contextos regulatórios que devem: i) possui um tipo de toxicidade (aguda, crônica etc.) definido, onde esta definição é determinante para o mecanismo de toxicidade interpretado;

ii) algoritmo não ambíguo, capazes de minimizar o risco de dúvidas geradas em sua utilização; iii) domínio de aplicabilidade definido, evitando a utilização e extrapolação inadequada; iv) medida apropriada de ajuste do modelo aos dados e v) interpretação mecanística, quando possível

(PÉREZ SANTÍN et al., 2021).

Um termo frequentemente utilizado nos estudos de previsão e toxicidade é o termo *in silico*. Este, é uma alusão aos termos *in vivo* e *in vitro*, que no contexto toxicológico, são consideradas metodologias de testes tradicionais.

Segundo Takata et al. (2020), o termo *in silico* é utilizado para descrever metodologias baseadas em uma abordagem computacional para previsão do comportamento toxicológico (qualitativo ou quantitativo) de determinada substância química, utilizando bases de dados contendo resultados experimentais reais, obtidos por métodos tradicionais de avaliação toxicológica.

Metodologias *in silico* para previsão de toxicidade incluem diversas técnicas tais como: aprendizado de máquina (*machine learning*), métodos SAR (*Structure-Activity-Relationship*) e aplicações *expert-based* (WANG; GOODMAN; ALLEN, 2021).

De acordo com Idakwo et al. (2019), os métodos SAR são baseados em modelos estatísticos/matemáticos e utilizados para estabelecer uma relação aproximada entre uma propriedade biológica de um composto químico e propriedades físico-químicas e/ou estruturais, por meio de seus grupos químicos funcionais. A previsão baseia-se na premissa de que moléculas químicas com estruturas e grupos funcionais similares exibem atividade biológica semelhante.

Os métodos SAR podem ser qualitativos ou quantitativos, neste último caso conhecidos como métodos QSAR (*Quantitative Structure-Activity-Relationship*). É importante ressaltar, que os métodos SAR não excluem a necessidade de aplicação de técnicas de *machine learning*. Para a aplicação de técnicas computacionais de análise de dados, os grupos funcionais das moléculas estudadas são codificados como descritores moleculares. Estes descritores, atrelados à estrutura química da molécula, são utilizados como variáveis independentes na etapa de modelagem, enquanto a variável dependente pode ser um valor numérico, tal como a concentração letal (CL) (nos casos de QSAR) ou uma variável categórica, tal como os rótulos tóxico e não tóxico (IDAKWO et al., 2019).

A modelagem computacional por métodos SAR para previsão de toxicidade é amplamente descrita na literatura, onde 60% dos estudos utilizaram esta abordagem para a previsão de toxicidade. Os demais 40% utilizam outras abordagens que não estão diretamente ligadas a relação estrutura – atividade.

Devido à premissa do método QSAR (relação estrutura – atividade quantitativa), comumente este tipo de modelo é utilizado para a modelagem toxicológica de substâncias

químicas isoladas (puras), não tratando de misturas, uma vez que a estrutura da molécula é considerada para determinação dos códigos descritores moleculares. Portanto, a literatura carece de relatos que tratem da previsão da toxicidade ou do comportamento toxicológico de misturas. Esta carência pode estar relacionada a diversos motivos, dentre eles os efeitos de sinergismo, potenciação e antagonismo derivado da interação entre as espécies químicas frente ao organismo teste, além da própria natureza generalista do teste de toxicidade aguda.

Dentre os relatos aqui analisados, apenas um estudo aborda a previsão toxicidade de misturas. Trata-se do estudo de Duan et al. (2020) que utilizou algoritmos de aprendizado de máquina para prever a toxicidade de misturas de três substâncias químicas utilizando a técnica IJP (*Ink Jet Printing*) e dados de toxicidade por luminescência.

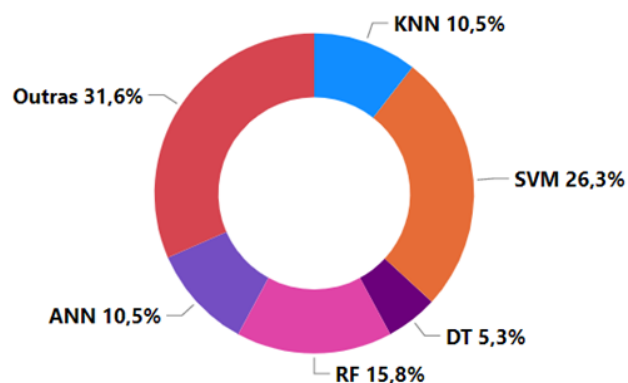


Figura 12: Percentual de técnicas utilizadas em relação ao total de técnicas apresentadas nos estudos.

Fonte: Autores.

Neste método, diversas concentrações de cada componente de uma mistura ternária foram submetidas a testes de toxicidade por luminescência e a resposta colorimétrica foi impressa em placas que, aliadas às concentrações conhecidas dos componentes químicos, serviram de base de dados para a previsão da toxicidade, sendo os algoritmos kNN (*k-Nearest Neighbors*) e SVM (*Support Vector Machine*) os que forneceram modelos de melhor performance quando comparados aos modelos de regressão linear (DUAN et al., 2020).

No que diz respeito às ferramentas e técnicas utilizadas para a previsão e toxicidade nos estudos analisados, foi observada uma diversidade de métodos e técnicas na busca pelo melhor modelo de aderência aos dados contidos nas diversas bases de dados utilizadas nos estudos. De uma forma geral, as etapas para a construção de um modelo, descritas nos estudos aqui analisados são: i) pré-processamento e padronização; ii) codificação dos

compostos químicos; iii) modelagem, avaliação e calibração do modelo e iv) definição do domínio de aplicabilidade. Esta sequência de etapas foi utilizada em diversos estudos, principalmente aqueles com foco na avaliação de modelos, tais como os estudos de Morger et al. (2021), Shadrin et al. (2020), Takata et al. (2020), Castillo-Garrit et al. (2017), Mishra, Fei, Huan (2013) e Yin, et al. (2011).

3.2.1 Pré-processamento e padronização dos dados

Na etapa de pré-processamento parte-se do princípio de que os dados de testes in vivo ou in vitro estão sujeitos a diversas fontes de erro e ruído e, sabendo-se que o poder de previsão de um modelo é tão melhor quanto melhor for a qualidade dos dados contidos na base de dados, é necessário um pré-processamento para que as fontes de erro e ruídos sejam minimizadas ou eliminadas (IDAKWO et al., 2019).

A verificação e o processamento dos dados brutos, nesta etapa é de extrema importância para a geração e modelos confiáveis. Muitas vezes, a verificação a qualidade dos dados é negligenciada o que pode resultar no desenvolvimento de modelos ruins. Em alguns casos, eles podem até representar bem os dados, mas não refletem bons resultados quando submetidos a um novo conjunto de dados de teste (WANG; GOODMAN; ALLEN, 2021).

Segundo Idakwo et al. (2019), a qualidade dos dados utilizados para o treinamento de um modelo é mais importante do que a escolha do algoritmo a ser utilizado. Entretanto, em alguns trabalhos, o método utilizado para limpeza e padronização dos dados não é claro.

Castillo-Garrit et al. (2017) afirmam que a utilização de atributos não necessários pode confundir o sistema e estes devem ser eliminados na etapa de pré-processamento.

Segundo Wang, Goodman, Allen (2021), a base de dados normalmente é processada de forma a se adequar para a aplicação dos algoritmos de aprendizado de máquina ou outros. Tarefas comuns nesta etapa incluem a remoção de valores em branco, inválidos ou considerados, por experiência própria ou de estudos anteriores, desnecessários. Além disto, problemas relacionados à bases de dados desbalanceadas devem ser mitigados nesta etapa e existem alguns métodos para esta mitigação e o mais comum deles é a reamostragem.

O desbalanceio do conjunto de dados ocorre quando, dentro deste conjunto, há um número significativamente diferente entre compostos ativos (tóxicos) e não ativos (não tóxicos), frente a um mesmo organismo teste e nas

mesmas condições de estudo, o que é frequente em bases de dados toxicológicas, uma vez que são analisados mais compostos não tóxicos do que tóxicos (TANG et al., 2019).

Idakwo et al. (2019) que mostrou que uma possível solução para problemas de desbalanceio é a reamostragem. Morger et al. (2021) utilizou análise conformacional para avaliar a calibração de um modelo. O estudo verificou que, devido ao desbalanceio dos dados na base de dados, o fator de calibração foi negativamente afetado e que isto pode ser melhorado utilizando-se bases de dados maiores ou contendo um maior número de observações de compostos tóxicos.

Esta questão do desbalanceio foi também observada por Tang et al. (2019) que concluiu que para a obtenção de modelos de alta performance, são necessários dados balanceados e suficientes, o que geralmente não ocorre.

Nos métodos de previsão QSAR, devem também ser retirados dos dados: sais, misturas, compostos inorgânicos ou organometálicos e demais compostos iônicos. Isto porque a relação estrutura-atividade compreende bem o comportamento biológico de moléculas orgânicas (YIN et al., 2015).

De todos os relatos aqui analisados, apenas o estudo de Takata et al. (2020) traz um modelo capaz de lidar com compostos orgânicos com múltiplos grupos funcionais, compostos inorgânicos e compostos iônicos, contudo, a técnica de modelagem foi desenvolvida pelos autores, não sendo, portanto, uma técnica comumente utilizada para previsão de toxicidade.

3.2.2 Codificação dos compostos químicos

Para o desenvolvimento de modelos QSAR, no que diz respeito à codificação dos compostos químicos com base em seus grupos funcionais, existem algumas formas de codificação frequentemente utilizadas e comumente descritas na literatura, tais como: descritores moleculares e fingerprints moleculares. Fingerprints moleculares, em geral, representam as moléculas por meio de rótulos (labels) binários em que cada parte (bit) representa fragmentos ou subestruturas da molécula. Já os descritores moleculares incluem características tais como contagem atômica, solubilidade, entre outros (WANG; GOODMAN; ALLEN, 2021, YANG et al., 2020; IDAKWO et al., 2019; CASTILLO-GARRIT et al., 2017; MISHRA; FEI; HUAN, 2013; SHERHOD et al., 2014).

3.2.3 Modelagem, avaliação e calibração do

modelo

Na etapa de modelagem, o modelo propriamente dito é desenvolvido por meio da utilização de técnicas capazes de extrair padrões nos dados de forma a possibilitar a previsão de resposta a um novo conjunto de dados e assim, prever a toxicidade de novos compostos. A confiabilidade do modelo recém-gerado é então avaliada e o modelo pode ser calibrado e reavaliado inúmeras vezes até que os melhores resultados da avaliação sejam obtidos (IDAKWO et al., 2019).

Técnicas de aprendizado de máquina são comumente utilizadas para a análise dos dados de previsão ou determinação de padrões toxicológicos de substâncias químicas, sendo a abordagem QSAR ou outra. A Tabela 02 apresenta as técnicas utilizadas em cada um dos estudos, bem como a base de dados utilizada.

Segundo Wang, Goodman, Allen (2021), o aprendizado de máquina é um campo recentemente utilizado na química computacional com inúmeras aplicações, sobretudo na descoberta de novas drogas, quimioinformática e previsão de toxicidade.

Ainda de acordo com Wang, Goodman, Allen (2021), na literatura estão descritos diversos algoritmos utilizados nesta previsão, sendo os mais comuns: kNN (k-Nearest Neighbors), SVM (Support Vector Machine), Naïve Bayes (NB), Árvores de Decisão (Decision Trees – DT), Random Forest (RF) e Redes Neurais (NN), podendo ser Redes Neurais Artificiais (Artificial Neural Network – ANN) ou Deep Neural Network (DNN). Técnicas que possuem maior facilidade de utilização, entendimento e generalização, são mais amplamente utilizadas.

A Figura 12 apresenta os modelos/algoritmos utilizados nos trabalhos analisados neste estudo e colabora para responder a questão de pesquisa QEP2. Nota-se, portanto, o predomínio de SVM e RF para a previsão de toxicidade nos estudos aqui analisados, o que corrobora os dados encontrados na revisão de Wang, Goodman, Allen (2021).

A calibração do modelo está relacionada à sua confiabilidade, ou seja, o quão bem o modelo proposto representa e/ou descreve os dados do conjunto de dados. A confiabilidade do modelo pode ser mensurada de diversas formas, a depender do modelo e do que está sendo previsto. Para os modelos QSAR, os parâmetros comuns de avaliação de confiabilidade são: sensibilidade, especificidade, acurácia e acurácia balanceada.

As formas de determinação de cada um destes parâmetros estão descritas no estudo de Tang et al, 2019. Outras formas de determinar à adesão do modelo aos

dados e calibrá-los também são apresentadas no mesmo estudo, sendo a acurácia balanceada mais utilizada para modelos QSAR, pois estes frequentemente apresentam dados desbalanceados e modelos binários, comumente de classificação.

Além desta, é corriqueira a utilização da área abaixo da curva ROC (Receiving Operation Curve), chamada de AUC (area under curve) que se trata da área abaixo da curva plotada pela sensibilidade em relação à taxa de falsos positivos, equivalente a uma especificidade (TANG et al., 2019).

Shadrin et al. (2020), mostraram que SVM e RF tiveram maiores correlações com os dados reais de fitotoxicidade que outros modelos, sendo que SVM apresentou $R^2 = 0,8$ e RMSE cerca de 11 a 12, enquanto ANN apresentou uma RMSE 11,05. Quando se trata de modelos de regressão, os parâmetros utilizados mais frequentemente são o coeficiente de determinação R^2 e o erro quadrático médio (RMSE) (IDAKWO et al., 2019).

Uma das principais ocorrências que devem ser avaliadas nesta etapa é o overfitting, que ocorre quando o modelo aprende tão bem com os dados durante a etapa de treino que acaba por memorizá-los, e quando, ao prever novos dados não utilizados no conjunto de dados de treino, o modelo performa mal ou pior que o esperado (WANG; GOODMAN; ALLEN, 2021).

Em contrapartida, os resultados de previsão utilizando-se o conjunto de dados de treino são muito precisos. Portanto, overfitting é um problema de alta importância no desenvolvimento de modelos por métodos computacionais e algoritmos mais complexos são considerados mais propensos a esta ocorrência. Por isto, frequentemente é necessário o uso de técnicas de regularização ou validação externa para limitar os efeitos do overfitting, porém isto não estabelece o quão aplicável o modelo é a novos dados. Para isto, é necessária a determinação e boa documentação do domínio de aplicabilidade do modelo (WANG; GOODMAN; ALLEN, 2021).

3.2.3.1 SVM – Support Vector Machine

SVM são classificadores de margem máxima, os quais minimizam o erro da classificação empírica e maximizam a margem geométrica que separa os dados. Normalmente resultam em um espaço de atributos no qual os dados são linearmente separados por um vetor, até mesmo quando o espaço inicial não é linear. São bons em reconhecimento de padrões, generalizam bem e lidam bem com dados de diversas dimensões (WANG; GOODMAN; ALLEN, 2021).

Yin, et al. (2011) afirmam que o SVM foi desenvolvido

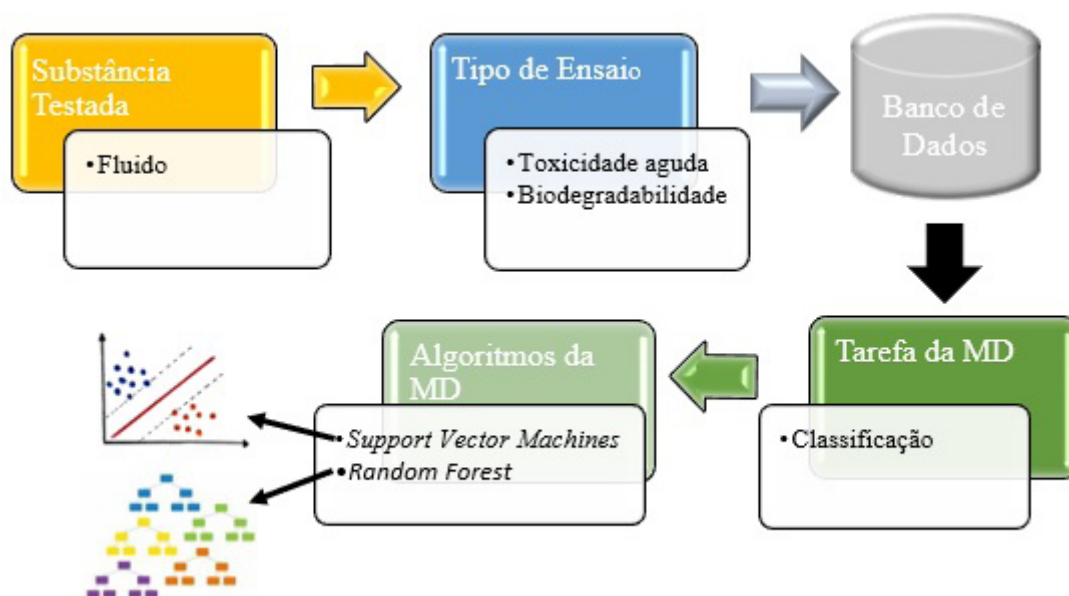


Figura 13: Técnica de Mineração de Dados aplicada a ecotoxicidade de fluidos de perfuração
Fonte: Autores.

para reconhecimento de padrões, sendo aplicado para analisar a linearidade entre os dados e, para dados não linearmente separados, amostras não linearmente separadas de um espaço de baixa dimensão são convertidas em um vetor que as separa linearmente em um espaço vetorial de alta dimensão, o que permite uma análise linear em um espaço de atributos não linearmente separados. Além disto, SVM é baseado na teoria da minimização de risco estrutural para construir o hiperplano de segmentação ótimo no espaço dos atributos.

De acordo com Fan et al. (2021), SVM é um algoritmo muito utilizado em previsão de toxicidade uma vez que o modelo gerado normalmente permite que a toxicidade conhecida de um composto químico pode ser utilizada para prever a toxicidade do mesmo composto frente a outras espécies não testadas. Além disto, oferece vantagens quando se trata de modelos de regressão com dados limitados, mesmo para dados toxicológicos nos quais o efeito tóxico observado em diferentes espécies por diferentes substâncias químicas, mesmo que semelhantes, apresenta uma relação não linear.

Por este motivo, Shadrin et al. (2020) afirmam que SVM é considerado um algoritmo promissor e eficiente no que tange à interpretação de dados altamente não lineares e dimensionais corriqueiramente utilizados em ciências ambientais, concluindo que a melhor previsão de fitotoxicidade, utilizando uma base de dados própria, foi encontrada utilizando-se SVMs.

3.2.3.2 RF - Random Forest

Random Forest (RF) é um método baseado em árvores de decisão (decision tree ou DT). Uma árvore de decisão trata-se de um método de avaliação em árvore constituído por raízes, nós e folhas, sendo que cada nó possui um único caminho desde a raiz e nele é tomada alguma decisão, pré-estabelecida de acordo com o conjunto de dados e o que se deseja obter como resposta, dependendo, portanto, de regras que por sua vez dependem dos parâmetros da base de dados (WANG; GOODMAN; ALLEN, 2021).

Em uma DT, começando pela raiz, a árvore se divide em dois ou mais ramos que se dividem em dois ou mais ramos até que em uma folha seja tomada a decisão final, o que, no caso de previsão e toxicidade, pode ser um rótulo: tóxico ou não tóxico (IDAKWO et al., 2019).

Mishra, Fei, Huan (2013) mostraram que ao reduzir o tamanho das DTs e manter apenas as que contribuem para uma correta tomada de decisão em um modelo RF, pode-se obter melhores performances. Além disto, afirmam que em previsão de toxicidade por RF, um dos desafios é o elevado número de atributos (descritores moleculares ou fingerprints moleculares) para descrever uma única molécula, o que causa uma alta dispersão no espaço de dados dos atributos, quando da utilização de abordagens QSAR.

RF é um método que opera pela construção de múltiplas DTs. A ideia básica de um método RF utilizado para previsão de toxicidade QSAR é a combinação de DTs desenvolvidas utilizando-se um conjunto de dados de descritores ou fingerprints moleculares para treinamento. Nesta etapa, os dados de treino são selecionados

aleatoriamente e isto faz com que as DTs geradas sejam muito diversas e as previsões mais generalizáveis.

Duan et al. (2020), ao analisar os dados de uma fonte própria utilizando regressão linear, kNN, SVM, RF e outros, concluíram que o de melhor aderência aos dados foi o RF. Este estudo é o único aqui analisado que se refere à previsão do efeito tóxico de misturas de substâncias químicas e, por ser um método composto, o RF pode ter performado melhor nesta previsão, tendo em vista a complexidade de previsão do comportamento toxicológico de misturas.

3.2.4. Determinação do domínio de aplicabilidade do modelo

Por fim, após a criação e determinação da acurácia do modelo, deve-se determinar e bem documentar o seu domínio de aplicabilidade. Conforme já mencionado, dentre os 5 propósitos propostos pela OCDE para o uso de técnicas computacionais em contextos regulatórios, está a necessidade de determinação de um domínio de aplicabilidade (PÉREZ SANTÍN et al., 2021).

Este fato é corroborado por Wang, Goodman, Allen (2021) que afirmam que a determinação e documentação do domínio de aplicabilidade é uma prática comum em metodologias QSAR e que este domínio tem sido relatado como um elemento chave na previsão de toxicidade in silico, tendo em vista que tais modelos podem subsidiar tomadas de decisão ou mesmo legislações. A utilização equivocada inclui a previsão, onde o dado de entrada do modelo não está contido no domínio de aplicabilidade.

De acordo com Yang et al. (2020), há diversas estratégias de determinação do domínio de aplicabilidade de modelos. Entretanto, cada uma delas tem suas próprias vantagens e desvantagens e para modelos que utilizam a metodologia QSAR, técnicas que combinem múltiplos métodos, tais como RF, apresentam domínios de aplicabilidade mais estruturados.

4. DISCUSSÃO

No que tange à ecotoxicologia de fluidos de perfuração, a maioria dos estudos analisados utiliza o fluido como amostra para a determinação da ecotoxicidade, assim como exige a legislação brasileira (não limitada a este parâmetro) e diversas outras ferramentas regulatórias ao redor do mundo. A maior parte dos estudos considerou também a toxicidade aguda como parâmetro principal de avaliação, contudo, estão amplamente documentados

estudos que avaliam a biodegradabilidade e toxicidade crônica.

A toxicidade de um fluido de perfuração é determinada pelos seus componentes e por suas propriedades físico-químicas. Contudo, alguns componentes dos fluidos de perfuração não influenciam positiva ou negativamente em sua toxicidade.

Seja por razões éticas, econômicas ou sociais, a utilização de organismos vivos para a realização de testes toxicológicos tradicionais in vitro ou in vivo vem sendo questionada. Neste sentido, a toxicologia computacional, por meios de métodos in silico é uma área de estudo que vem ganhando notoriedade por possibilitar a minimização da realização de testes tradicionais, sobretudo para fins regulatórios.

Assim, o estabelecimento de metodologias in silico capazes de determinar os padrões toxicológicos de fluidos de perfuração comumente utilizados na indústria de óleo e gás pode viabilizar a redução do número de testes realizados com organismos vivos e possibilitar a geração de conhecimento científico, atrelado à descoberta dos padrões toxicológicos destes fluidos, com base em sua composição e propriedades físico-químicas.

Um dos métodos mais amplamente descritos na literatura para a previsão de toxicidade de compostos químicos é o QSAR, que se baseia na relação estrutura – atividade de substâncias químicas. Este método necessita da codificação da estrutura molecular das substâncias químicas como dado de entrada para a geração dos modelos.

Existem bases de dados específicas que contêm estes dados codificados em descritores ou fingerprints moleculares. O modelo preditivo depende da quantidade de descritores ou fingerprint que uma molécula possui, assim como das propriedades físico-químicas desta substância, sendo utilizado para prever a toxicidade de novas moléculas, pela codificação destas, sem a necessidade de repetir o teste tradicional.

Modelos matemáticos, com base em dados contendo resultados de testes tradicionais já realizados, podem prever a toxicidade de fluidos de perfuração ainda não testados, ou mesmo fornecer informações relevantes no que diz respeito ao comportamento toxicológico destes fluidos. Contudo, é imprescindível que os dados contidos no conjunto de dados sejam confiáveis e reproduzíveis, pois quanto melhor for sua qualidade, melhor a previsão e a acurácia do modelo gerado.

Diversas técnicas computacionais vêm sendo utilizadas para a previsão de toxicidade, sejam elas mais simples ou mais complexas, na busca pelo conhecimento

e previsão de alta qualidade, com domínio de aplicabilidade bem definido de forma a possibilitar o uso desta previsão para questões regulatórias, tomadas de decisão, análises de risco, dentre outros.

Dos estudos analisados, apenas um, Duan et al. (2020), estudou o comportamento toxicológico de misturas, caso dos fluidos de perfuração. Para este cenário, parece mais interessante e relevante a aplicação de técnicas de mineração de dados para a determinação de padrões e geração de conhecimento do que para a previsão da toxicidade propriamente dita (quanti ou qualitativa), devido à complexidade das misturas.

Seja o método de previsão com base na relação estrutura-atividade (SAR) ou não, as técnicas mais utilizadas para a previsão de toxicidade são SVM e RF.

Em síntese, pode-se responder as duas questões específicas de pesquisa propostas (QEP1 e QEP2), assim como propor uma resposta para a questão de pesquisa geral (QP), já que não foi possível obter uma resposta para esta questão diretamente da literatura, conforme explicado anteriormente.

QEP1 – Como é feita a Ecotoxicologia de fluidos de perfuração?

No que tange à ecotoxicologia de fluidos de perfuração, a maioria dos estudos analisados utiliza o fluido como amostra para a determinação da ecotoxicidade, e como parâmetros principais de avaliação nos fluidos foram utilizados a toxicidade aguda e a biodegradabilidade.

QEP2 – Como a Mineração de Dados é utilizada para previsão da toxicidade?

Em relação a previsão de toxicidade, os algoritmos de MD mais utilizados foram SVM e RF. De acordo com os estudos selecionados, esses algoritmos possibilitaram à aplicação da tarefa de MD conhecida como classificação.

QP – Quais são as técnicas computacionais mais utilizadas para a previsão e/ou conhecimento do comportamento toxicológico de fluidos de perfuração, conhecida a sua composição e suas propriedades físico-químicas?

Fazendo uma análise das respostas obtidas para as duas questões específicas de pesquisa (QEP1 e QEP2) por meio da literatura foi possível propor uma resposta capaz de atender a questão de pesquisa geral (QP)

Conforme mostra a Figura 13, este trabalho propõe responder a questão geral de pesquisa proposta por

meio da utilização de fluido a partir de ensaio de toxicidade aguda e biodegradabilidade, onde os resultados obtidos são armazenados na base de dados.

Ainda conforme a Figura 13, para a previsão e/ou conhecimento do comportamento toxicológico dos fluidos são aplicadas técnicas computacionais de Mineração de Dados, também conhecida como Machine Learning, utilizando a tarefa de classificação por meio dos algoritmos Support Vector Machines (SVM) e Random Forest (RF).

5. CONCLUSÕES

A toxicologia computacional vem sendo muito estudada nos últimos anos, o que mostra a relevância desta abordagem na previsão do comportamento toxicológico de compostos químicos.

No que se refere à ecotoxicidade de fluidos de perfuração, a literatura apresenta diversos estudos que relacionam os componentes destes fluidos com sua ecotoxicidade, contudo não há estudos que utilizem técnicas ou ferramentas computacionais para a previsão do comportamento toxicológico de fluidos de perfuração, dentro dos critérios de busca adotados no presente estudo.

Portanto, este trabalho focou em responder separadamente as questões específicas de pesquisa QEP1 e QEP2, de modo que uma análise das respostas obtidas proporcionasse uma resposta a questão geral de pesquisa, a qual não foi possível de se responder diretamente da literatura analisada.

5.1. Agenda de Trabalhos Futuros

A partir desta revisão sistemática foi possível observar a carência de estudos primários referentes à previsão do comportamento toxicológico de misturas. Como caso particular, de fluidos de perfuração que são amplamente utilizados nas perfurações de poços de óleo e gás. Portanto, há uma lacuna de pesquisa que pode ser suprida por estudos futuros com este foco. Em trabalhos futuros, é promissor o desenvolvimento de processos de extração de conhecimento em bases de dados ou mesmo de previsão do comportamento toxicológico de misturas. Pode-se propor como trabalhos futuros a aplicação da MD por meio da tarefa de classificação para extração de conhecimento e previsão do comportamento toxicológico de fluidos de perfuração. Propõe-se também aplicar as demais tarefas de MD (regressão, associação e clusterização) e comparar os resultados obtidos.

Tais estudos podem ser utilizados por operadoras, no

sentido de aprofundar o conhecimento a respeito dos fluidos de perfuração utilizados atrelado à necessidade de atender aos critérios das agências reguladoras e órgãos ambientais e influenciar seus fornecedores neste sentido.

Os fabricantes e fornecedores dos produtos químicos que compõem o fluido de perfuração podem usufruir do conhecimento gerado nestes estudos no intuito de atuar no desenvolvimento de produtos mais ambientalmente amigáveis, substituindo aqueles que agregam maior toxicidade.

Além destes, os próprios órgãos ambientais e agências reguladoras também são beneficiados com estes estudos, com ênfase em aprimorar seu entendimento relacionado ao impacto ambiental causado pela utilização e descarte dos fluidos de perfuração, bem como proporcionar a minimização ou extinção da realização de testes in vivo ou in vitro ou mesmo direcionar novas ações na busca constante da proteção ambiental.

REFERÊNCIAS

ASLAN, J. F.; **Estudos ecotoxicológicos na perfuração de poços de Petróleo Marítimos com ênfase na atividade de cimentação**. Dissertação de Mestrado, Instituto Federal de Ciência e Tecnologia Fluminense. Macaé, 2018. 59f.

ASSOCIAÇÃO BRASILEIRA DE NORMAS TÉCNICAS (ABNT). **NBR 15308**: ecotoxicologia aquática – toxicidade aguda – método de ensaio com misídeos (Crustacea). Rio de Janeiro, 2017.

ASSOCIAÇÃO BRASILEIRA DE NORMAS TÉCNICAS (ABNT). **NBR 15469**: ecotoxicologia – coleta, preservação e preparo de amostras. Rio de Janeiro, 2015.

BAKER HUGHES. **Drilling Fluids - Reference Manual**. Fonte: Research Gate Website. 775 p, 2016. Disponível em: <https://www.researchgate.net/file.PostFileLoader.html?id=5733bd33615e2775127a46b7&assetKey=AS:360698208112640@1463008562446>. Acesso em 01 de outubro de 2022.

BAKHTYAR, S; GAGNON, M. Toxicity assessment of individual ingredients of synthetic-based drilling muds (SBMs). **Environmental Monitoring and Assessment**, v. 184, n. 9, p. 5311 – 5325, 2012.

BRASIL. Ministério do Meio Ambiente. **Despacho Nº**

5540547/2019, SEI 5533803 de 29 de julho de 2019. Define diretrizes para uso e descarte de fluidos de perfuração e cascalhos. Brasília: Instituto Nacional do Meio Ambiente e Recursos Naturais Renováveis, 2019. Disponível em: https://sei.ibama.gov.br/controlador_externo.php?acao=usuario_externo_logar&id_orgao_acesso_externo=0. Acesso em 01 de outubro de 2022.

CASTILLO-GARRIT, J. et al. Machine learning-based models to predict modes of toxic action of phenols to *Tetrahymena pyriformis*. **SAR and QSAR in Environmental Research**, v. 28, n. 9, p. 735-747, 2017.

CHEN, W. et al. Life cycle toxicity assessment on deep-brine well drilling. **Journal of Cleaner Production**, v. 112, p. 326 – 332, 2016.

DUAN, Qiannan et al. Machine learning for mixture toxicity analysis based on high-throughput printing technology. **Talanta**, v. 207, p. 120299, 2020.

DURGUT, I. et. al. Dynamic modeling of environmental risk associated with drilling. **Marine Pollution Bulletin**, v. 99, n. 1 – 2, p. 240 – 249, 2015.

FAN, Juntal et al. Prediction of chemical reproductive toxicity to aquatic species using a machine learning model: An application in an ecological risk assessment of the Yangtze River, China. **Science of the Total Environment**, v. 796, p. 148901, 2021.

FERRARI, D. G.; SILVA, L. N. **Introdução a mineração de dados**. Saraiva Educação, 2017.

GAGNON, M. M.; BAKHTYAR, S. Induction of fish biomarkers by synthetic-based drilling muds. **PLoS ONE**, v. 8, n. 7, p. e69489, 2013.

GE, W. F. et. al. A “dual protection” drilling fluid system and its application. **Petroleum Science and Technology**, v. 30, n. 12, p. 1274 – 1284, 2012.

IDAKWO, Gabriel et al. A review on machine learning methods for in silico toxicity prediction. **Journal of Environmental Science and Health - Part C**, v. 36, n.4, p. 169-191, 2019.

LIRA, V. F. et. al. Effects of barium and cadmium on the population development of the marine nematode

Rhabditis (Pellioditis) marina. **Marine Environmental Research**, 72, n. 4, p. 151 – 159, 2011.

LORENZETT, C. D. C.; TELOKEN, A. **Estudo Comparativo entre os algoritmos de Mineração de Dados Random Forest e J48 na tomada de Decisão**. Universidade de Cruz Alta (UNICRUZ), 2016.

MA, Chao Y.; BUONTEMPO, Frances V.; WANG, Xue Z. Inductive data mining: Automatic generation of decision trees from data for QSAR modelling and process historical data analysis. **Journal of Modelling, Identification and Control**, v. 12, n. 1/2, p. 101-106, 2011.

MISHRA, Meenakshi; FEI, Hongliang; HUAN, Jun. Computational prediction of toxicity. **International Journal of Data Mining and Bioinformatics**, v. 8, n. 3, p. 338-348, 2013.

MORGER, Andrea et al. Assessing the calibration in toxicological in vitro models with conformal prediction. **Journal of Cheminformatics**, v. 13, n. 1, 2021.

OGELEKA, D. F.; TUDARARO-AHEROBO, L. E. Short-term toxicity of oil-based drilling fluid to the brackish-water shrimp *Palaemonetes africanus*. **African Journal of Aquatic Science**, v. 36, n.1, p. 109 – 112, 2011.

PAGE, M.J.; MCKENZIE, J. E.; BOSSUYT, P.M.; BOUTRON I.; HOFFMAN, T.C.; MULROW, C. D. et al. The PRISMA 2020 statement: an updated guideline for reporting systematic reviews. **BMJ**, v. 372, n. 71, 2021.

PÉREZ SANTÍN, Efrén et al. Toxicity prediction based on artificial intelligence: A multidisciplinary overview. **Wiley Interdisciplinary Reviews Computational Molecular Science**, v. 99, n. 11, 2021.

QU, Y. et. al. Influence of various hydrocarbon groups on the effectiveness and environmental characteristics of anti-collapse agent for drilling fluids. **Chemistry and Technology of Fuels and Oils**, v. 56, n 3, p. 420 – 428, 2020.

SADIQ, R. et. al. Marine water quality assessment of synthetic-based drilling waste discharges. **International Journal of Environmental Studies**, v. 60, n.4, p. 313 – 323, 2003.

SHADRIN, Dmitrii et al. Artificial intelligence models to predict acute phytotoxicity in petroleum contaminated soils. **Ecotoxicology and Environmental Safety**, v. 194, p. 110410, 2020.

SHERHOD, Richard et. al. Emerging pattern mining to aid toxicological knowledge discovery. **Journal of Chemical Information and Modeling**, v. 54, n. 7, p. 1864-1879, 2014.

STELIGA, T; ULIASZ, M. Spent drilling muds management and natural environment protection. **Mineral Resources Management**, v. 30, n. 2, p. 135 – 159, 2014.

SIL, A. et. al. Toxicity characteristics of drilling mud and its effect on aquatic fish populations. **Journal of Hazardous, Toxic, and Radioactive Waste**, v. 16, n.1, p. 51 – 57, 2012.

SUN, Y. et. al. Study on a New Environmentally Friendly Synthetic Fluid for Preparing Synthetic-Based Drilling Fluid. **Frontiers in Chemistry**, v. 8:539690, 2020.

TAKATA, Michiyoshi et al. Predicting the acute ecotoxicity of chemical substances by machine learning using graph theory. **Chemosphere**, v. 238, p. 124604, 2020.

TANG, Weihao et al. Deep learning for predicting toxicity of chemicals: a mini review. **Journal of Environmental Science and Health - Part C**, v. 36, n. 4, p. 252-271, 2019.

VEIGA, L. F. **Estudo da toxicidade marinha de fluidos de perfuração de poços de óleo e gás**. Dissertação de Mestrado, Instituto de Biologia Marinha, Universidade Federal Fluminense, Niterói, 1998. 120 f.

VEIGA, L. F. **Avaliação de risco ecológico dos descartes da atividade de perfuração de poços de óleo e gás em ambientes marinhos**. Tese de Doutorado, Coordenação dos Programas de Pós-graduação em Engenharia, Programa de Engenharia Civil, Universidade Federal do Rio de Janeiro, Rio de Janeiro, 2010. 253f.

VINCENT-AKPU, I. F.; ALLISON, M. E.; SIKOKI, F. D. Survival and gill morphology of different life stages of *Tilapia guineensis* exposed to the drilling fluid XP-07. **Ciência Rural**, v. 40, n.3, p. 611 – 616, 2010.

WANG, Marcus W. H; GOODMAN, Jonathan, M.; ALLEN, Timothy E. H. Machine Learning in Predictive Toxicology: Recent Applications and Future Directions for Classification Models. **Chemical Research in Toxicology**, v. 34, n. 2, p. 217-239, 2021.

WITTEN, I. H.; FRANK, E.; HALL, M. A. **Data Mining: Practical Machine Learning Tools and Techniques**. 4th edition ed. Burlington, MA: Morgan Kaufmann, 2016.

XIE, S. X. et. al. Study and application of green throwing drilling fluid. **Petroleum science and technology**, v. 30, n. 5, p. 443 – 452, 2012.

YANG, Hongbin et al. Computational Approaches to Identify Structural Alerts and Their Applications in Environmental Toxicology and Drug Discovery. **Chemical Research in Toxicology**, v. 33, n. 6, p. 1312-1322, 2020.

YIN, Yongmin et al. Quantitative Regression Models for the Prediction of Chemical Properties by an Efficient Workflow. **Molecular Informatics**, v. 34, n. 10, p. 679-688, 2015.

AUTORES

ORCID: <https://orcid.org/0000-0002-8286-5192>

LEOMIR SAMUEL TORMEN REIS | Universidade Federal de Viçosa, MG - Brasil | Mestrando em Engenharia Ambiental, Instituto Federal Fluminense, Macaé, RJ - Brasil | Correspondência para: Rua Itaipu, 91/101. Praia do Pecado, Macaé - RJ, 27920-120 | email: leomir.tormen@gmail.com

ORCID: <https://orcid.org/0000-0002-9130-8902>

VICTOR BARBOSA SARAIVA, Mestre em Ciências Biológicas, Doutor em Ciências Biológicas. | Instituto Federal Fluminense, Cabo Frio, RJ - Brasil | Correspondência para: Rua do Guriri, 261, Quadra G Casa 25A. Però, Cabo Frio, RJ. CEP: 28922-370 | email: victor.saraiva@gsuite.iff.edu.br

ORCID: <https://orcid.org/0000-0002-5994-6840>

SIMONE VASCONCELOS SILVA, Doutora em Computação. | Instituto Federal Fluminense, Campos dos Goytacazes, RJ - Brasil. | Correspondência para: R. Virgílio de Paula, 35 - Tarcísio Miranda, Campos dos Goytacazes - RJ, 28020-130. | email: simonevsinfo@gmail.com

COMO CITAR ESTE ARTIGO

REIS, Leomir Samuel Tormen; SARAIVA, Victor Barbosa; SILVA, Simone Vasconcelos. **Mineração de Dados e Previsão da Ecotoxicidade de Fluidos de Perfuração – Uma Revisão Sistemática**. *MIX Sustentável*, v. 9, n. 2, p. xx-xx, 2023. ISSN 2447-3073. Disponível em: <<http://www.nexos.ufsc.br/index.php/mixsustenta-vel>>. Acesso em: _/_/_. doi: <<https://doi.org/10.29183/2447-3073.MIX2023.v9.n2.xx-xx>>.

SUBMETIDO EM: 12/07/2021

ACEITO EM: 05/10/2022

PUBLICADO EM: 31/03/2023

EDITORES RESPONSÁVEIS: José Manuel Couceiro Barosa e Paulo Cesar Machado Ferroli.

Registro da contribuição de autoria:

Taxonomia CRediT (<http://credit.niso.org/>)

LSTR: conceituação, curadoria dos dados, análise formal, investigação, metodologia, visualização, escrita - rascunho original.

VBS: conceituação, análise formal, investigação, metodologia, administração de projetos, supervisão, validação, escrita - revisão & edição.

SVS: conceituação, análise formal, investigação, metodologia, administração de projetos, supervisão, validação, escrita - revisão & edição.

Declaração de conflito: nada foi declarado.