



UNILAB

**UNIVERSIDADE DA INTEGRAÇÃO INTERNACIONAL DA
LUSOFONIA AFRO-BRASILEIRA
INSTITUTO DE ENGENHARIAS E DESENVOLVIMENTO SUSTENTÁVEL
CURSO DE ENGENHARIA DE COMPUTAÇÃO**

MISAEAL BESSA SALES

**UMA REVISÃO ABRANGENTE SOBRE O PAPEL TRANSFORMADOR DA
INTELIGÊNCIA ARTIFICIAL EM ESTUDOS COM ENZIMAS E SUA
RELEVÂNCIA NA INDÚSTRIA**

ACARAPE – CE
2023

MISAEEL BESSA SALES

**UMA REVISÃO ABRANGENTE SOBRE O PAPEL TRANSFORMADOR DA
INTELIGÊNCIA ARTIFICIAL EM ESTUDOS COM ENZIMAS E SUA
RELEVÂNCIA NA INDÚSTRIA**

Trabalho de conclusão de curso apresentado ao curso de Engenharia de Computação da Universidade da Integração Internacional da Lusofonia Afro-Brasileira como requisito parcial à obtenção do título de Engenheiro de Computação.

Orientador: Prof. Dr. José Cleiton Sousa dos Santos.

Universidade da Integração Internacional da Lusofonia Afro-Brasileira
Sistema de Bibliotecas da UNILAB
Catalogação de Publicação na Fonte.

Sales, Misael Bessa.

S163r

Uma revisão abrangente sobre o papel transformador da inteligência artificial em estudos com enzimas e sua relevância na indústria / Misael Bessa Sales. - Redenção, 2023. 94f: il.

Monografia - Curso de Engenharia Da Computação, Instituto De Engenharias E Desenvolvimento Sustentável, Universidade da Integração Internacional da Lusofonia Afro-Brasileira, Redenção, 2023.

Orientador: Prof. Dr. José Cleiton Sousa dos Santos.

1. Enzimas. 2. Inteligência artificial. 3. Machine learning.
4. Deep learning. 5. Análise bibliométrica avançada. I. Título

CE/UF/Dsibiuni

CDD 612.0151

MISAEEL BESSA SALES

UMA REVISÃO ABRANGENTE SOBRE O PAPEL TRANSFORMADOR DA
INTELIGÊNCIA ARTIFICIAL EM ESTUDOS COM ENZIMAS E SUA RELEVÂNCIA
NA INDÚSTRIA

Trabalho de conclusão de curso apresentado ao curso de Engenharia de Computação da Universidade da Integração Internacional da Lusofonia Afro-Brasileira como requisito parcial à obtenção do título de Engenheiro de Computação.

Orientador: Prof. Dr. José Cleiton Sousa dos Santos

Aprovado em: 30/11/2023

BANCA EXAMINADORA

Prof. Dr. José Cleiton Sousa dos Santos
Universidade da Integração Internacional da Lusofonia Afro-Brasileira

Prof^a. Dr^a. Maria Cristiane Martins de Souza
Universidade da Integração Internacional da Lusofonia Afro-Brasileira

Prof^a. Dr^a. Márcia Roberta Falcão de Farias
Universidade da Integração Internacional da Lusofonia Afro-Brasileira

AGRADECIMENTOS

Aos meus pais Messias e Rosimeire pelo apoio, cuidado e incentivo imprescindíveis durante toda a minha trajetória acadêmica. Eu os amo demais e isso tudo tem se deve a minha relação familiar e todo meu sucesso são para eles.

Aos meus amigos e colegas de curso Calebe, Cleiton, Gregory e Yuri por sempre torcerem por mim e terem tido uma participação assídua durante todo o processo de conhecimento me ajudando e me apoiando sempre quando precisava.

A Vítor e Graça que são pessoas que sempre posso contar como apoio e me dão a sua positividade para seguir em frente e servindo de grande exemplo para eu continuar lutando pelos meus sonhos e meu sucesso.

Aos meus amigos de longa data Milton e Guilherme que são pilares essenciais dentro do processo de aprendizado de vida e me deram grande suporte emocional para conseguir ultrapassar barreiras e dificuldades que aparecem na minha vida.

Ao meu orientador Prof. Dr. José Cleiton Sousa pela confiança, que durante 2 anos de trabalhos contínuos sempre esteve a disposição para me ajudar e serviu de pilar inspirativo para a busca de conhecimento, evolução profissional e pessoal.

A UNILAB e ao IEDS por disponibilizar profissionais excelentes que permitiram grande conhecimento aprofundado sobre os assuntos de estudo.

Ao GENEZ (Grupo de Engenharia Enzimática) pelo apoio e compartilhamento de conhecimento.

A todas as pessoas que contribuíram direta ou indiretamente para esta conquista, minha eterna gratidão a todos!

Resumo

Os últimos anos de desenvolvimentos no campo da Inteligência Artificial (IA) foram possíveis graças ao aumento da disponibilidade de grandes quantidades de dados, que são resultados da capacidade de transformar digitalmente a informação e das recentes evoluções de poder computacional. No estudo com enzimas as técnicas de IA podem contribuir com diversos campos como o design de enzimas, previsão de funções e estabilidade catalítica, descoberta de novas enzimas e otimização de processos. Neste estudo foram exploradas aplicações de IA nos diversos modos de estudo com enzimas, demonstrando dados de tendência de uso através de técnicas de revisão bibliométrica que quantificam subcampos de investigação e disponibiliza análise da literatura científica mundial. As abordagens de Machine Learning (ML) se mostraram muito presentes principalmente no subcampo da predição de processos enzimáticos que englobam verificação de melhor espaço conformacional do sítio ativo da molécula ou a reação de uma enzima com um possível alvo. Para a indústria farmacêutica, as IA têm papel muito importante na produção de drogas mais eficientes e eficazes. Esta área de estudo se mostrou muito promissora e com alto potencial futuro em uma grande variedade de abordagens.

Palavras-chave: Enzimas; Inteligência Artificial; Machine Learning; Deep Learning; Análise Bibliométrica Avançada.

Abstract

Recent years of developments in the field of Artificial Intelligence (AI) have been made possible thanks to the increased availability of large amounts of data, which are the result of the ability to digitally transform information and recent developments in computing power. In the study of enzymes, AI techniques can contribute to various fields such as enzyme design, predicting catalytic functions and stability, discovering new enzymes and optimizing processes. This study explored AI applications in the various ways of studying enzymes, demonstrating trend data through bibliometric review techniques that quantify research subfields and provide an analysis of the world's scientific literature. Machine Learning (ML) approaches have shown themselves to be very present, especially in the subfield of predicting enzymatic processes that include checking the best conformational space for the molecule's active site or the reaction of an enzyme with a possible target. For the pharmaceutical industry, AI plays a very important role in the production of more efficient and effective drugs. This area of study has shown great promise and high future potential in a wide variety of approaches.

Keywords: Enzymes; Artificial Intelligence; Machine Learning; Deep Learning; Advanced Bibliometric Analysis.

LISTA DE FIGURAS

Figura 1. Diagrama dos subconjuntos da Inteligência Artificial.	13
Figura 2. Estrutura que representa os critérios de pesquisa e análise.	18
Figura 3. Crescimento das publicações e citações de 2010 até ao início de 2023. Esta figura foi construída utilizando o Microsoft Excel sob dados coletados na plataforma WoSCC através da análise de resultados de busca.	24
Figura 4. Mapa de visualização em rede relacionado com a colaboração entre autores; o número de citações define o tamanho de cada item. A espessura das linhas que ligam cada autor a outro indica a acumulação de co-autoria, e os agrupamentos divididos em cores diferentes ilustram os grupos de autores com um elevado nível de colaboração.	29
Figura 5. a) Mapa de visualização da rede de colaboração entre países com pelo menos 5 artigos publicados. b) Mapa de visualização de sobreposição onde as cores da escala representam a distribuição periódica da relevância por país. c) Representação do número de artigos publicados por país no período analisado e a rede de colaboração. d) Gráfico que mostra uma roda de dependência entre os países, onde cada ponto é composto por várias relações ponderadas com outros pontos através de sua força de ligação. Esse modelo de gráfico é muito utilizado para entender o fluxo de dados e pode ser uma excelente forma de ilustrar relações entre dados. e) Gráfico que mostra o nível de citação dos países.	36
Figura 6. a) Mapa de visualização da rede relacionada à colaboração entre organizações com pelo menos 10 documentos. A espessura das linhas que ligam as duas organizações indica fortemente o acúmulo de coautorias, e as cores divididas em clusters ilustram grupos de instituições com alto grau de colaboração. b) Geocodificação das organizações registradas nos 1292 artigos analisados.	38
Figura 7. Relação de três campos da pesquisa em estudo (Autor, País do Autor e Tópico de Pesquisa).	39
Figura 8. Distribuição das áreas de pesquisa definidas na base de dados em investigação.	44
Figura 9. Funcionamento genérico de um algoritmo de machine learning no tratamento computacional de uma enzima.	46
Figura 10. Tendências emergentes na área de pesquisa.	61
Figura 11. Mapa de visualização das palavras-chave dos 5 países com mais publicações.	63
Figura 12. (A) Número de patentes registradas nos países com maior destaque nas pesquisas. (B) Número de patentes registradas durante o período de 2014 a 2023.	64

LISTA DE TABELAS

Tabela 1. Classificação das 10 principais revistas, países, afiliações e autores com base no número de publicações e citações no domínio da Inteligência Artificial em estudos com enzimas.....	32
Tabela 2. Os artigos mais citados na área de investigação.	43
Tabela 3. Métodos de aprendizado de máquina no campo da predição das abordagens enzimáticas.	47
Tabela 4. As 24 principais palavras-chave do autor na área de Inteligência Artificial em estudos enzimáticos.....	60

SUMÁRIO

1. INTRODUÇÃO.	10
2.1 METODOLOGIA.	17
2.1 Fonte dos Dados.	17
2.2 Métodos Bibliométricos.	20
2.3 Questões de Pesquisa.	24
3 UMA REVISÃO ABRANGENTE SOBRE O PAPEL TRANSFORMADOR DA INTELIGÊNCIA ARTIFICIAL EM ESTUDOS COM ENZIMAS E SUA RELEVÂNCIA NA INDÚSTRIA.	24
3.1 Análise Bibliométrica.....	24
3.1.1 Relações de coautoria.	27
3.1.2 As revistas científicas e o estudo dos parâmetros bibliométricos.	29
3.1.3 A distribuição por país e instituição.	34
3.1.4 Publicações mais citadas e áreas de investigação.	39
3.2 Overview das principais utilizações da IA nos processos de engenharia enzimática. .	44
3.2.1 Previsão de enzimas utilizando Machine Learning.	46
3.2.2 Aplicações de Inteligência Artificial no design de enzimas.....	51
3.2.3 Otimização de Processos.....	52
3.2.4 Descoberta de novas enzimas.....	53
3.2.5 Estabilidade enzimática.....	53
3.2.6 Automação de Laboratórios.	55
3.2.7 Obstáculos e oportunidades no uso das IAs.	56
3.2.8 Aplicação na indústria farmacêutica.	57
4. TÓPICOS DE PESQUISA EMERGENTES.	58
4.1 Análise Quantitativa das palavras-chave frequentes.....	58
4.2 Áreas de Pesquisa.....	61
5 UMA REVISÃO DAS PATENTES DE IA NO CAMPO DA ENGENHARIA ENZIMÁTICA.	63
6. CONCLUSÃO	65
7. REFERÊNCIAS	67

1. INTRODUÇÃO

Recentemente a Inteligência Artificial (IA) obteve uma grande ascensão que, se devidamente aproveitado, pode oferecer um promissor uso em muitos setores de aplicação em variados campos. (Barredo Arrieta *et al.*, 2020; Dow *et al.*, 2022; Jaremko *et al.*, 2019; Kutyauro; Rushambwa; Chiwazi, 2023; Sheikh; Fann, 2019; Yang *et al.*, 2021). O objetivo da IA é simular a inteligência humana em máquinas que são programadas para se comportar de uma determinada maneira e pensar de forma semelhante aos humanos (Chakraborty *et al.*, 2021; Deig; Kanwar; Thompson, 2019; Haleem; Javaid; Khan, 2019; Jimma, 2023; Luo, Jiefeng *et al.*, 2023; Muthukrishnan *et al.*, 2020; Niel; Bastard, 2019; Olabi *et al.*, 2023; Salehi; Burgueño, 2018; Sarno *et al.*, 2023). Os últimos anos de desenvolvimentos no campo da IA foram possíveis graças ao aumento da disponibilidade de grandes quantidades de dados, que são resultados da capacidade de transformar digitalmente a informação e das recentes evoluções de poder computacional (Balsano *et al.*, 2022; Fan, 2023). De maneira mais precisa, as IAs são programas algorítmicos de software computacional que realizam a simulação das funções neurais (redes neurais) e as interações (neurotransmissão sináptica) entre as camadas neurais progressivas no cérebro humano (Catania; Nicolitz, 2018). Os processos de utilização da IA incluem: a possibilidade de fazer com que as máquinas resolvam problemas com alta dificuldade em uma velocidade muito mais rápida do que o cérebro humano levaria para chegar em uma solução para o mesmo problema (Kar; Choudhary; Singh, 2022); análise de big data para analisar e compreender tendências e fazer previsões de cenários futuros da forma mais otimizada possível com o menor desperdício de tempo e esforço (Arfanuzzaman, 2021; Duan; Edwards; Dwivedi, 2019; Jiao *et al.*, 2023; Mainali; Park, 2023; Yahyaoui *et al.*, 2023; Zhang *et al.*, 2021); e é gerado uma enorme economia no custo operacional (Agwu *et al.*, 2018).

Dentro do campo da inteligência artificial existem os subcampos de aplicações "Machine Learning" e "Deep Learning" assim como mostrado na figura 1. O Aprendizado de

Máquina (ML) é o processo de utilização de dados e algoritmos para simular o procedimento de aprendizagem computacional e aprimorar a precisão dos sistemas (Mystridis et al., 2022; Okafor et al., 2023; Zhong, Xu et al., 2022; Zhong; Salehi; Johnson, 2022). O aprendizado profundo (DL) é um subconjunto do ML que utiliza redes neurais profundas (DNNs) (Brendel et al., 2022; Pacal et al., 2020; Tian et al., 2023; Zheng et al., 2021). O ML e especialmente as técnicas de DL ganharam muita popularidade nos últimos anos em diversos campos de pesquisa, como processamento de linguagem natural (Barber e Colab., 2021; Kaczmarek e Colab., 2022; RIBELLES e colab., 2021), processamento de imagem (Aljabri et al., 2022; Bhatt et al., 2022; Zhang; Kwong; Wang, 2020) e reconhecimento de fala (Kaur; Rattan, 2023; Kumar, L. Ashok et al., 2022; Mulfari et al., 2021). Além disso, foram desenvolvidas técnicas para modelar fenômenos químicos e biológicos com maiores níveis de complexidade através dessas ferramentas (Xiao; Verkhivker; Tao, 2022) como o docking molecular ligante-proteína para descobrimento de fármacos (Crampon et al., 2022a); digestão anaeróbica utilizada na recuperação de recursos como produtos químicos de plataforma de alto valor e digerido rico em nutrientes e bioenergia de diversos resíduos orgânicos (Khanal; Tarafdar; You, 2023); Transcrição e tradução genética (Chow et al., 2020) e a previsão de metástase (Albaradei et al., 2021). O aprendizado profundo realiza a permissão para que modelos computacionais estruturados por diversas camadas de processamento possam aprender representações de dados com muitos níveis de abstração (Guarascio; Manco; Ritacco, 2019). Esses métodos melhoraram drasticamente o estado da arte em processos de reconhecimento de voz, reconhecimento visual e detecção de objetos, dentre outros domínios, como descoberta de drogas e genômica (Cao et al., 2018). O aprendizado profundo explora e descobre estruturas complexas em grandes aglomerados de dados usando o algoritmo de retropropagação para indicar como uma máquina deverá modificar seus parâmetros internos utilizados no cálculo da representação em cada camada a partir da representação na camada anterior (recursividade) (LECUN, 2015).

As aplicações de IA ultrapassam gradualmente as paredes seguras dos laboratórios de pesquisa e se amarram a nossa vida cotidiana. Muitas aplicações de IA do mundo real, por exemplo, na medicina (Bruckert; Finzel; Schmid, 2020), na direção autônoma (Wehner *et al.*, 2022) e no desenvolvimento de produtos industriais (Schramm; Pieper; Vogl, 2022; Wehner; Kertel; Wewerka, 2023), são críticas para a segurança (Schramm; Wehner; Schmid, 2023). A IA pode ser classificada em três tipos: IA Focada, IA Generalizada e IA Superinteligente. A IA Focada, também chamada de IA Fraca ou Estreita, consiste em algoritmos especializados em resolver problemas em um campo e/ou um problema específico. Neste caso os sistemas guardam uma grande quantidade de informações e os algoritmos são capazes de realizar tarefas complexas, entretanto sempre focadas no objetivo para o qual foram construídos. Os Sistemas Especialistas e Sistemas de Recomendação são exemplos de sistemas de IA focada (Ludermir, 2021). Na IA Generalizada, também conhecida como IA Forte, os algoritmos desenvolvidos têm capacidade semelhante ou superior a humanos em várias tarefas e, em geral, os algoritmos usam noções de Aprendizado de Máquina como estrutura. Em algumas tarefas os algoritmos têm desempenho semelhante aos humanos, por exemplo, em Visão Computacional (Peng *et al.*, 2023). O nível atual da IA é de IA Generalizada. Na IA Superinteligente, os algoritmos são consideravelmente mais capazes que humanos em praticamente todas as tarefas. Ainda não existem sistemas com IA Superinteligente e não se tem conhecimento se existirão sistemas mais inteligentes que os humanos desenvolvidos com técnicas de IA (Schneider, 2017).

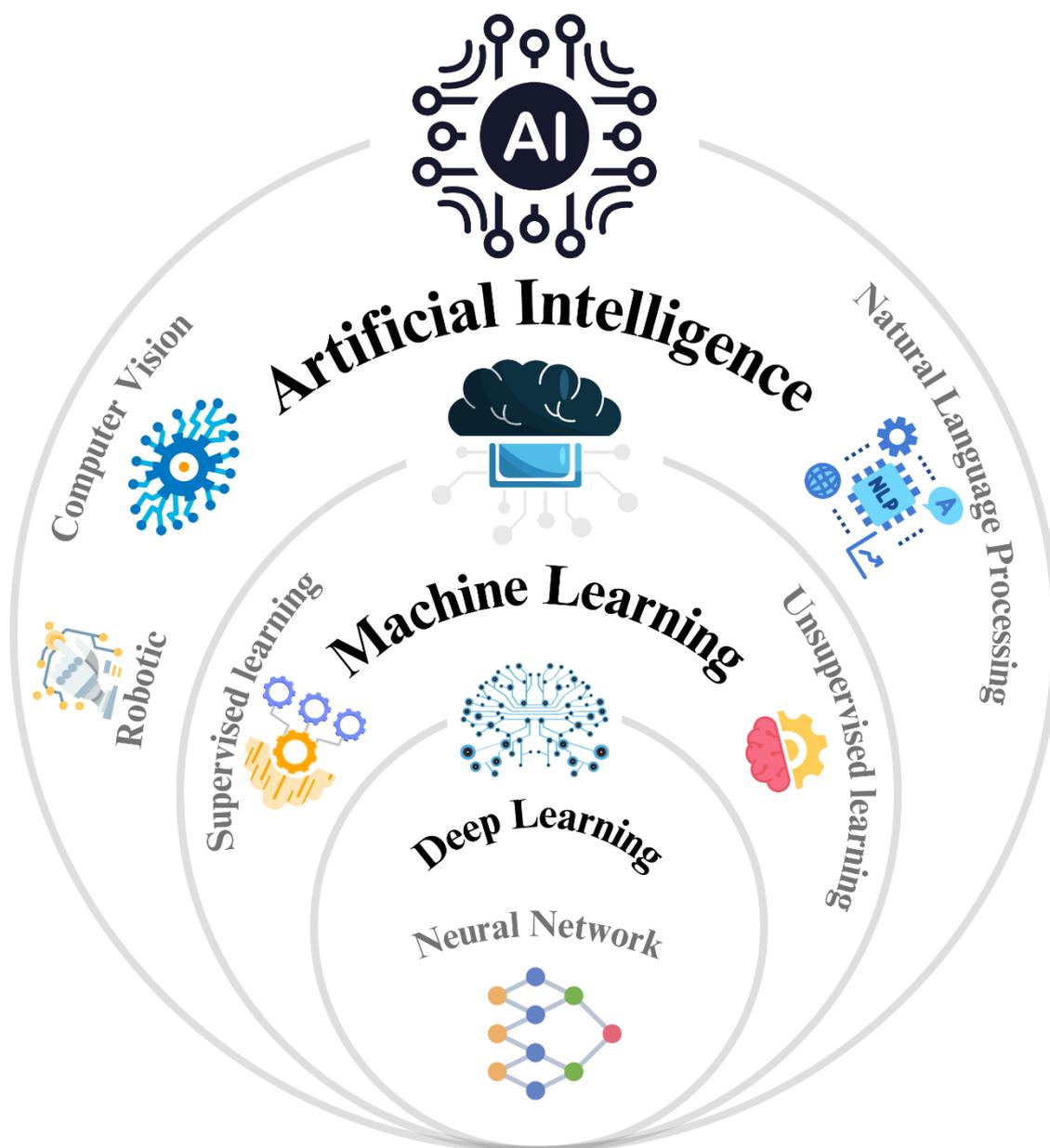


Figura 1: Diagrama dos subconjuntos da Inteligência Artificial.

No estudo com enzimas as técnicas de inteligência artificial podem contribuir com diversos campos como a evolução dirigida de enzimas no qual são acumuladas mutações benéficas com a função desejada por rodadas iterativas de mutagênese e triagem (Jang *et al.*, 2022a; Li; Dong; Reetz, 2019a), modelo de aprendizado profundo para desenvolver um modelo de Rede Neural Convolutacional (CNN) tendo como referência um conjunto de dados de enzimas Proteinase do SARS-CoV utilizado para prever a atividade anti-SARS-CoV em compostos desconhecidos (Kumari; Subbarao, 2021) (um tópico de pesquisa que obteve grande ascensão

nos últimos anos devido a pandemia do COVID 19) e até mesmo técnicas para realizar predições sobre características e funções as quais enzimas são a principal ferramenta de experimento ou estão diretamente interligadas no estudo (Cong *et al.*, 2009; Fernandes *et al.*, 2020; Memon; Aamir Khan; Naveed, 2020; Plonka *et al.*, 2021). Pode-se ainda citar usos concretos das técnicas de IA neste campo a avaliação e previsão da atividade enzimática do solo (Shahare *et al.*, 2023; Zhang; Lin; Chen, 2023a), evolução dirigida de enzimas seletivas (Li; Dong; Reetz, 2019b), classificação de enzimas (Nallapareddy; Dwivedula, 2021) e Modelagem simultânea de produção de enzimas (Günay *et al.*, 2008).

As enzimas são classificadas como catalisadores sustentáveis da natureza. Elas são biodegradáveis, biocompatíveis e são derivados de recursos renováveis. Os processos enzimáticos são conduzidos em condições amenas (aproximadamente a temperatura ambiente, pressão atmosférica e pH fisiológico) em água, com taxas elevadas e seletividades. Em adição a isso, a utilização de enzimas normalmente elimina a necessidade de ativação e/ou proteção do grupo funcional, disponibilizando rotas sintéticas mais econômicas, que resultam em menos resíduos e são mais eficientes energeticamente do que as sínteses orgânicas convencionais (Sheldon; van Pelt, 2013). As enzimas são catalisadores bioquímicos de extrema importância para diversos processos biológicos metabólicos, como crescimento e divisão celular. Aplicado a condições de reação distintas, elas são baseadas principalmente na aceleração da formação de ligações químicas, reduzindo a energia necessária para o estado de transição da reação entre o produto e o reagente (de Sousa *et al.*, 2023). Essas biomoléculas têm sido fonte de diversos estudos em muitas áreas de pesquisa como recursos energéticos sustentáveis (Alexandre *et al.*, 2022; Cavalcante *et al.*, 2021, 2022; Ferreira Mota *et al.*, 2022), no acoplamento com outros compostos para desenvolvimento de biossensores (Silva *et al.*, 2022; Souza *et al.*, 2022) e outras aplicações relacionadas a biocatálise (Nunes *et al.*, 2021; Ortiz *et al.*, 2019; Velasco-Lozano; Rocha-Martin; Santos, 2022). O atual crescimento de métodos experimentais de alto

rendimento, incluindo triagem automatizada e sequenciamento de próxima geração, sobrecarregou a área da biologia molecular com big data, o que obrigou os pesquisadores a repensarem suas abordagens simultâneas consoante à engenharia de enzimas. A Inteligência Artificial e o Aprendizado de Máquina têm alto potencial para revolucionar a engenharia enzimática inteligente sem a necessidade notável de um entendimento total do sistema molecular subjacente (Singh *et al.*, 2021).

Para análise sistemática da abordagem deste artigo foi utilizado um método de revisão bibliométrica avançada para realizar uma verificação mais precisa de aspectos gerais que vão compor a estrutura científica literária. Este tipo de análise é uma metodologia popular e rigorosa para investigar altas quantidades de dados científicos (Abdelwahab *et al.*, 2023). Isso permite a visualização da evolução de um tópico de pesquisa importante, além de revelar informações essenciais sobre as áreas que estão se desenvolvendo rapidamente neste campo (Tamala *et al.*, 2022). A análise bibliométrica avançada se utiliza de um conjunto de dados com grandes quantidades de documentos, que podem descrever dados específicos de pesquisas, como o número de ocorrências de determinadas palavras-chave, o volume de citações e/ou publicações e relações de cooperação entre grupos de autores (Sales *et al.*, 2022). Entretanto, para um melhor entendimento dos aspectos é necessário, além das informações quantitativas, investigar sobre os assuntos emergentes de maneira mais sistemática, adotando uma análise manual. Essa metodologia é muito eficaz para observar características de pesquisa muito importantes quando se trata de tratamento de muitos dados (Rodrigues *et al.*, 2023). Enquanto as revisões convencionais resumem em geral a literatura existente com base nas opiniões dos pesquisadores, que podem estar expostas a vários tipos de vieses, a análise bibliométrica acrescenta rigor por causa de uma avaliação quantitativa e estruturada das publicações (Rejeb *et al.*, 2022). O principal benefício da bibliometria, em comparação com os métodos tradicionais de revisão, é que ela disponibiliza a construção de mapas científicos com altos

índices de precisão, imparciais, abrangentes e rigorosos. Tais mapas constituem visualizações avaliativas de interações entre entidades científicas com base nas ligações entre esses dispositivos (Zhang *et al.*, 2020). Os benefícios do mapeamento científico também incluem o potencial para analisar grandes conjuntos de dados (Big Data) e desenvolver visualizações interpretáveis, memoráveis e ricas (Balakrishnan; Haney; Meuer, 2016). Diante disso, a aplicação da bibliometria oferece uma abordagem oportuna, imparcial e visual para rastrear o desenvolvimento e analisar a estrutura intelectual de um determinado campo de pesquisa. Não só facilita retrospectivas de investigação histórica, mas também permite descobrir tendências e pontos críticos em domínios de conhecimento a partir de perspectivas macro e micro (Li *et al.*, 2017). Neste trabalho foram considerados artigos que no título, resumo ou palavras-chave estivessem relacionados a IA ou seus respectivos grandes sub-conjuntos (ML e DL), para obter-se uma quantidade equilibrada que pudesse ser processada sem muitos problemas e ao mesmo tempo fosse uma amostra significativa para análise.

Ainda, com base no que foi mostrado, este estudo abordou uma revisão atual e uma análise bibliométrica das pesquisas científicas sobre o uso das técnicas que utilizam as tecnologias de IA e sua modulação para diversos problemas e perspectivas nos estudos com enzimas. Esta revisão tem como foco a análise da progressão da pesquisa, evoluções e tendências dos métodos de aplicação do processamento computacional para esta temática. Esta revisão contribui para a descrição das tecnologias disponíveis que demonstrem potencial efetivo no estudo utilizando processos enzimáticos, disponibilizando a possibilidade de avaliar as variadas aplicações das IAs e explorar oportunidades inovadoras para uso das ferramentas em processos específicos. Portanto, este estudo efetivamente propõe pesquisas futuras e desenvolve o conhecimento sobre os novos rumos industriais do uso das inovações tecnológicas em inteligência artificial nas subáreas que envolvem os estudos de processos enzimáticos. Esse formato de investigação pode ser observado em vários artigos de análise avançada como o de

Sales (Sales *et al.*, 2022) e Rodrigues (Rodrigues *et al.*, 2023).

2. Metodologia

2.1. Fonte dos dados

Para o desenvolvimento deste trabalho foi utilizado o formato de pesquisa encontrado na plataforma de trabalhos científicos Web of Science - Core Collection (<https://www.webofscience.com/wos/woscc/basic-search>) utilizando-se dos operadores de pesquisas que permitem uma alta delimitação e especificidade de uma determinada abordagem de estudo. E com isso é disponibilizado, em formato de base de dados, um arquivo baixável contendo diversas informações sobre os artigos relacionados como título, ano de publicação, resumo, autores envolvidos na pesquisa e entre outros. Esses arquivos vão ser utilizados em um processamento computacional e a partir disso gerar muitas análises importantes sobre esse campo de estudo. O acesso ao site da Web of Science - Core Collection (WoSCC) foi possível devido a disponibilização de credenciais de login pela plataforma educacional CAPES PERIÓDICOS (<https://www-periodicos-capes-gov-br.ez373.periodicos.capes.gov.br/>, acesso em 19 de Fevereiro de 2023).

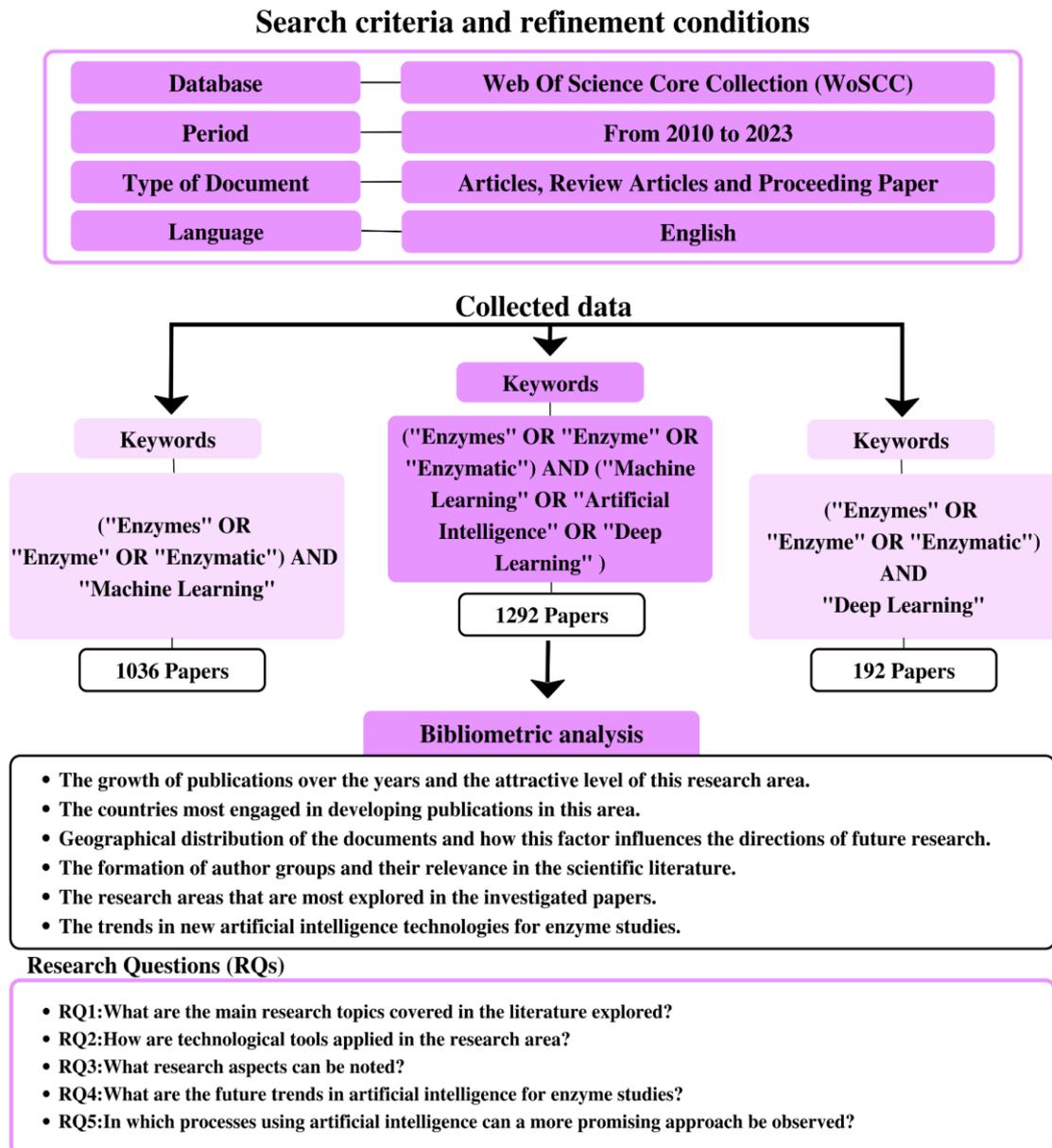


Figura 2: Estrutura que representa os critérios de pesquisa e análise.

A figura 2 ilustra a abordagem metodológica utilizada para modelar o presente estudo, identificando as características palavras-chave de busca, condições e critérios de refinamento, análise pós-processamento computacional e questões que serão respondidas no decorrer do trabalho. Para a realização do processo de busca foi definido o período de 2010 a 2023 no operador “Anos de Publicação”, isso para uma visualização abrangente tanto da clara curva de crescimento que pode ser visualizada na figura 3 como na delimitação de uma boa quantidade de artigos que contribuam para a economia do custo computacional disponível. Com

isso é possível obter uma amostragem ideal da temática em investigação para fácil visualização, exploração de suas características e avaliação quantitativa/qualitativa. Na pesquisa foi utilizado o operador “Tópicos” e as palavras-chave "Enzymes" OR "Enzyme" OR "Enzymatic" no primeiro campo de busca, e no segundo foi inserido “Artificial Intelligence” OR “Machine Learning” OR “Deep Learning” com o mesmo operador. A partir disso o processamento de busca resultou em 1292 artigos qualificados para a análise bibliométrica avançada. A utilização do operador ‘Tópicos’ vai informar à plataforma que deverá ser verificado a existência das palavras-chaves no título, resumo ou nas palavras-chave do autor. Para uma busca mais aprofundada dos conhecimentos científicos deste campo de pesquisa foram buscadas mais duas bases de dados que se utilizam da mesma delimitação de período, operadores e primeiro campo de pesquisa, modificando apenas a linha posterior. A segunda base de dados utilizou o termo “Machine Learning” e obteve 1036 papers. E, por fim, a terceira base utilizou-se da palavra-chave “Deep Learning” gerando apenas 192 documentos. Todas as bases de dados foram refinadas para que só existissem documentos do tipo Artigos, Artigos de Revisão e Artigos de Conferência e exclusivamente na linguagem inglês. A escolha desses operadores e termos de pesquisa se deu pela busca de uma base de dados que contivesse uma amostra de fácil processamento computacional e que disponibilizasse uma boa visualização e dados concisos sobre o assunto abordado. Além disso o período escolhido se deve pela avaliação de outros trabalhos que utilizam esta abordagem analítica de revisão avançada (Guimarães *et al.*, 2023; Huang *et al.*, 2023; Lee *et al.*, 2021; Zhang, Fangrong *et al.*, 2023), que definem um período de no mínimo de 10 anos ou mais para apresentar boas visualizações de gráficos e análises específicas como tabelas e figuras que necessitem de uma delimitação temporal. O período de 14 anos para avaliação também se deu pela curva gráfica explorada no banco de dados, onde após a pesquisa a WosCC disponibiliza uma prévia visualização da evolução durante os anos, tendo em vista isso, escolheu-se anos que a curva apresentasse um aumento significativo e

variante durante os anos, excluindo anos que não projetassem nenhum, ou poucos, artigos. A exclusividade da linguagem inglesa ocorreu devido ao processamento computacional dos programas, pois é possível que ao deixar dois termos que representam o mesmo sentido, mas são de linguagens diferentes, o algoritmo os trate como palavras diferentes, por tanto, para evitar impasses avaliativos, adotou-se essa característica.

2.2 Métodos Bibliométricos

A metodologia abordada neste artigo foi definida utilizando os artigos “*Sustainability and Challenges in Hydrogen Production: An Advanced Bibliometric Analysis*” publicado em 2023 pelo autor Catumba em 2023 (Catumba *et al.*, 2023) e “*Artificial Intelligence in Renewable Energy: A Comprehensive Bibliometric Analysis*” publicado pelo autor Zhang em 2022 (Zhang; Ling; Lin, 2022). Os autores têm a preocupação em adotar um formato de pesquisa que explore de forma mais abrangente possível e mais precisamente as características de estudo, utilizando-se do processamento computacional para realização da análise bibliométrica avançada. Além disso, explicam que esse método bibliométrico muito contribui para a análise dos rumos científicos emergentes nas áreas de investigação. A análise se dá principalmente por meio da base de dados com a maior quantidade de artigos que é utilizada no tratamento via softwares, mas as outras bases de dados servem, principalmente, para compreender os subcampos da inteligência artificial nos estudos com enzimas e prover informações importantes sobre eles.

A base de dados com 1036 artigos vai contribuir para a compreensão da utilização do Aprendizado de Máquina em diversos estudos que utilizam as enzimas, com a utilização de suas diversas ferramentas aplicáveis a diversas áreas dentro dessa investigação. E percebe-se que essa base representa grande parcela de todos os artigos coletados, mostrando sua grande importância. A terceira base com 192 artigos é caracterizada pela exploração do subcampo do Aprendizado Profundo que caracteriza um tópico mais específico para aplicações no estudo

com enzimas. A partir disso pode-se desenvolver um trabalho conciso sobre essas novas tecnologias no ramo da engenharia enzimática e suas diversas subdivisões de linha de pesquisa. Para realização do processamento computacional que este trabalho utiliza como principal meio de avaliação bibliométrica aprofundada, foram utilizados alguns softwares especializados nesse formato de pesquisa. Para a visualização da Clusterização de aspectos de pesquisa como a interconexão de países e organizações que mais publicam, índice de colaboração entre autores, principais áreas de pesquisas e palavras-chaves que identificam tendências de estudo foram utilizados o VOSviewer e o pacote Bibliometrix. O software VOSviewer foi construído no Centro de Estudos de Ciência e Tecnologia (CWTS) da Universidade de Leiden, Universidade de Leiden, Holanda. Este software vai ajudar a entender as estruturas correlacionais entre aspectos de extrema importância quando se trata de uma revisão bibliométrica, isso ajuda a explicar possíveis características de desenvolvimento científico que serão abordados neste trabalho. Bibliometrix é um conjunto de funcionalidades que compõe uma biblioteca que é executada dentro da linguagem de programação para estatísticas R, foi desenvolvido por Massimo Aria e Corrado Cuccurullo (Aria; Cuccurullo, 2017). Esta biblioteca contribuiu fortemente para a diversificação de visualizações que enriquecessem o trabalho, principalmente porque possui estruturas correlacionais diversas que vão complementar o VOSviewer. ArcGIS (versão 10.5 Redlands, Califórnia, EUA), ArcGIS é um conjunto de produtos desenvolvido pelo Environmental Systems Research Institute (Esri), Redlands, Califórnia, Estados Unidos. ArcGIS foi utilizado para realização do processo de geocodificação, que se trata da visualização de posição geográfica dos endereços dados traduzidos a pontos de localização (Haberman *et al.*, 2021). Para o processamento de dados na clusterização foram utilizadas métricas que possibilitassem a visualização facilitada, como o agrupamento de em média 30 nós em um mapa e cores distintas e destacadas para cada grupo.

2.3 Questões de pesquisa

Diante do processo metodológico utilizado, na aplicação da análise aprofundada algumas questões de pesquisa foram colocadas em evidência para ajudar na formulação do estudo. O desenvolvimento dessas questões se baseia em informações pertinentes que são colocados como norteamento para enriquecimento do artigo, quem por sua vez levou em conta trabalhos que utilizam o mesmo método como o de Shome (Shome *et al.*, 2023), Ranjbari(Ranjbari *et al.*, 2022), entre outros (Behera; Bala; Rana, 2023; Chiu *et al.*, 2023; Haddaji; Ayed; Fourati, 2022; Zarei; Khan; Abbassi, 2023). Essas questões de pesquisas são apresentadas a seguir:

- RQ1: Quais são os principais tópicos de investigação abordados na literatura explorada?

Entender os tópicos de investigação é uma das principais abordagens a se considerar devido a forma como a área de pesquisa se desenvolve durante o passar dos anos. Isso, pois há uma necessidade de explorar as características que podem ser modificadas ao longo do tempo, como interesse em uma nova subárea de pesquisa relacionada (Safdar *et al.*, 2023; Sales *et al.*, 2023). Essa questão ajuda a entender os rumos da literatura sobre as Ias que se envolvem com os estudos com enzimas.

- RQ2: Como são aplicadas as ferramentas tecnológicas na área da investigação?

O uso das diversas ferramentas que fazem parte das tecnologias de Inteligência Artificial é aplicado a diversos problemas que vão compor a área das enzimas (Levy *et al.*, 2023; Li *et al.*, 2018). Analisar a metodologia de aplicação delas é um fator fundamental para a compreensão abrangente do funcionamento geral desta investigação.

- RQ3: Que aspectos de investigação podem ser notados?

Existe uma variedade de aspectos pertinentes na avaliação da área de interesse, como localização geográfica, relações institucionais de cooperação, ligações individuais e em conjunto de colaboração no desenvolvimento de pesquisas, dentre outros. Muitos trabalhos vêm analisando essas características pois revela a formação da estrutura do campo investigativo

demonstrando como ele se comporta. O processo de desenvolvimento de uma área deixa diversos rastros para análise que vão ajudar a esclarecer pontos que uma inspeção simples não conseguiria sozinha contribuir de maneira significativa. Trabalhos como o de Ellegaard intitulado “*The bibliometric analysis of scholarly production: How great is the impact?*” (Ellegaard; Wallin, 2015) e o de Donthu nomeado “*How to conduct a bibliometric analysis: An overview and guidelines*” (Donthu *et al.*, 2021), que vão descrever de maneira clara esses fatos citados.

- RQ4: Quais são as tendências futuras em IA para estudos enzimáticos?

Como se trata de um trabalho que busca compreender os caminhos seguidos pelo campo de pesquisa em investigação, o processo tem como resultado uma visão das tendências futuras que satisfazem o interesse de construção de conhecimento científico. A análise detalhada mostra possibilidades evidentes em diversas subáreas e isso vai ajudar no desenvolvimento de novos projetos (Sales *et al.*, 2023).

- RQ5: Em que processos que utilizam inteligência artificial se pode observar uma abordagem mais promissora?

Esta análise tem um outro objetivo acoplado à construção da investigação avançada: o estudo das metodologias, ferramentas e processos. Isso se deve pelo fato de que são tecnologias associadas a formulação de técnicas metodológicas computacionais nos estudos de enzimas, por isso é necessário entender como um determinado procedimento é construído, quais linguagens de máquina são adotadas e quais estratégias de desenvolvimento utilizadas. Um exemplo disso são as atuais Redes Neurais Artificiais que são fortemente fonte de estudos em vários campos (Anitescu *et al.*, 2019; Araujo *et al.*, 2020; Berg; Nyström, 2018; Driscoll; de la Torre; Gomez-Ruiz, 2022; Ghobadian *et al.*, 2009; Komeilibirjandi *et al.*, 2020; Kotsiantis; Zaharakis; Pintelas, 2006; Li *et al.*, 2020; Mahdavi-Meymand; Sulisz, 2023; Moayedi *et al.*, 2020; Okuyucu; Kurt; Arcaklioglu, 2007; Ostad-Ali-Askari; Shayannejad; Ghorbanizadeh-

Kharazi, 2017; Pauletto; Dotto; Salau, 2020; Tuan Hoang *et al.*, 2021; Valipour; Banihabib; Behbahani, 2013; Wang *et al.*, 2017; Wu *et al.*, 2016; Yang *et al.*, 2018; Ying *et al.*, 2009; Yuan *et al.*, 2023).

3 UMA REVISÃO ABRANGENTE SOBRE O PAPEL TRANSFORMADOR DA INTELIGÊNCIA ARTIFICIAL EM ESTUDOS COM ENZIMAS E SUA RELEVÂNCIA NA INDÚSTRIA

3.1. Análise Bibliométrica

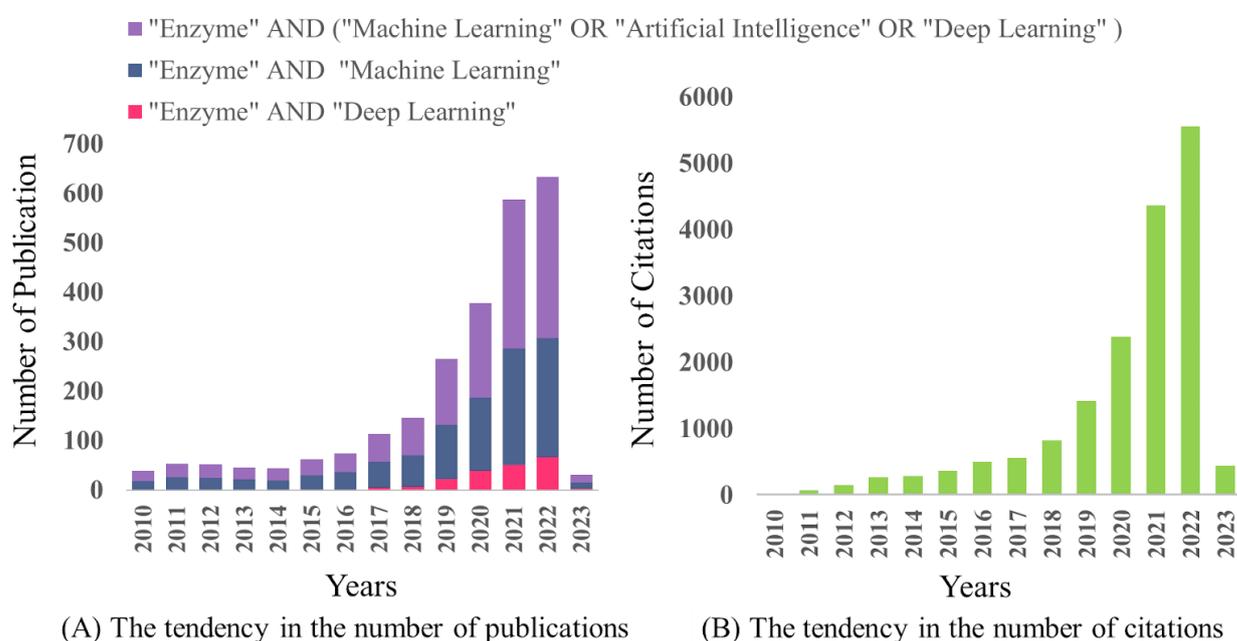


Figura 3: Crescimento das publicações e citações de 2010 até ao início de 2023. Esta figura foi construída utilizando o Microsoft Excel sob dados coletados na plataforma WoSCC através da análise de resultados de busca.

A partir da figura 3 pode-se compreender que a temática em estudo demonstra um grande crescimento tanto de quantidade de publicações (A) quanto de citações (B) durante o período especificado. Isso demonstra o aumento do interesse de exploração pelos pesquisadores. A figura 1 ajuda a compreender a relação entre as bases de dados demonstrada na parte A da figura 2, pois mostra que a Inteligência Artificial possui como subárea o Machine Learning que por sua vez possui o Deep Learning com subcampo. A partir desse diagrama que

a base de dados foi desenvolvida e manipulada para a construção deste trabalho. Os anos de 2021 e 2022 foram os anos que obtiveram um maior pico de quantidade de artigos publicados e de citações. 2022 teve 634 produções (maior base com 326, base mediana com 242 e menor base com 66) e um acúmulo de 5545 citações gerais. Esse número representa cerca de 49% da quantidade de artigos coletados para estudo e em comparação com o primeiro ano do período delimitado, há por volta de 16 vezes mais papers. Todas as bases de dados possuem a mesma tendência de crescimento anual que só se diferem pela quantidade absoluta de artigos publicados. Uma especificidade da figura 3 é a presença do ano de 2023 que apresenta poucos artigos, isso se deve pelo fato de que a coleta dos arquivos foi feita no início deste mesmo ano e por isso há poucos registros até o momento do processo de download. Em termos de impacto científico, a análise da figura 1 revela que a curva de crescimento durante o período em avaliação representa uma possível tendência que irá manter-se nos próximos anos (considerando a média de crescimento anual). Com isso, para este campo de pesquisa, isso configura um crescimento substancial que pode ser muito interessante para a sociedade científica, atraindo interesse de investigação.

A bibliometria explora as características fundamentais formais de domínios de conhecimento utilizando procedimentos matemáticos e estatísticos (Mora; Bolici; Deakin, 2017). A análise de rede de visualização utilizando instrumentos bibliométricos pode ser extremamente forte para compreender áreas temáticas estabelecidas e emergentes. Além disso, pode ajudar a identificar pesquisadores e os clusters de pesquisa revelando como as diversas áreas de pensamento podem ter surgido com base nos autores e aspectos organizacionais (Fahimnia; Sarkis; Davarzani, 2015). Identificar os pesquisadores mais influentes apresentados nos clusters prepara o terreno para conseguir determinar campos de estudo emergentes adicionais através da captura de tópicos mais recentes abordados por esses pesquisadores (Xu *et al.*, 2018). Consoante à isso, para avaliação procedural dos aspectos de pesquisa que serão

manipulados, serão considerados os seguintes métodos de processamento de dados:

- Análise do número de citações
 - Avaliar a relevância de cada tópico analisado e compreender o crescimento e como isso influencia o assunto abordado.
- Nível de relação de colaboração e coautoria
 - Entender como acontecem as relações e como elas são classificadas a partir de números como quantidade de citações acumuladas, quantidade de artigos publicados etc.
- Quantidade de ocorrências
 - Investigar quantas vezes uma determinada palavra aparece e entender o interesse investigativo sob ela.
- Análise de localização geográfica institucional
 - Verificar condições geográficas que influenciam as pesquisas, como por exemplo países que se interessam mais por subáreas que se relacionam com assuntos que pertencem a realidade deles (indústrias, educação, tecnologia etc.).

Todas essas análises revelam aspectos importantes que enriquecem o conhecimento sobre o campo de pesquisa abordado. A análise do número de citações é utilizada para ajudar a observar o crescimento científico em uma área específica em três diferentes níveis: micro, macro e meso (Derviş, 2020). A análise do nível de coautoria define a colaboração entre autores, podendo identificar vínculos tanto no âmbito intelectual quanto social. Essa análise pode ser útil na identificação inicial de grupos de pesquisa (Rodríguez-Soler; Uribe-Toril; De Pablo Valenciano, 2020). Na avaliação da quantidade de ocorrências, por exemplo de palavras-chave, nos permite analisar as palavras mais relevantes utilizadas nos dados (Owolabi; Sajjad, 2023). E por fim a análise de localização geográfica que é essencial para compreender as diferenças no investimento de variadas instituições em pesquisa em um determinado campo (Chen *et al.*,

2021). Cada uma das análises citadas engloba outros tipos de sub análises que vão ser mostradas mais detalhadamente nesta revisão.

3.1.1 Relações de coautoria

Uma primeira visão de extrema importância sobre os aspectos abordados na revisão bibliométrica é a relação de colaboração entre os diversos autores e grupos de autores existentes no meio acadêmico. As relações de coautoria revelam como está ocorrendo o desenvolvimento de artigos científicos para a área em análise. A figura 4 ilustra esta ligação e é perceptível o quão bem definido estão os grupos colaborativos e suas relações externas. É possível observar 5 cluster representados por cores diferentes para cada e sobretudo a força de relacionamento colaborativo destacado pelas linhas que os ligam e sua respectiva espessura. Um cluster é a representação esquemática de pontos que se relacionam através de linhas formando grupos que se diferenciam por um processo de similaridade, onde um cluster abriga um conjunto N de nós que se assemelham em algum aspecto definido previamente no programa. Ao analisar a figura é possível ver que o grupo que estão presentes os autores Li Y. e Zhang Y. é onde se concentra o maior número de autores relevantes tratados na investigação, e isso mostra que esse grupo permite uma maior diversificação de autores que contribuem com as pesquisas. É possível verificar a característica citada através de uma simples análise visual, onde no cluster informado está presente nós com maior tamanho. O tamanho desses nós são definidos pela quantidade de citações acumuladas por cada autor, revelando assim um nível de relevância. Consoante à isso, pode-se visualizar que existe um grupo (destacado em laranja) que é o oposto do citado anteriormente, mantendo suas relações de maneira restrita a um grupo pequeno de pesquisadores. Como esta é uma visualização criada a partir de autores com mais destaque, esse grupo menor enfrenta o problema de ter em seu grupo pesquisadores com índices de citação e publicação menores e acabaram não sendo considerados pelo algoritmo. Essa figura demonstra um fato interessante sobre as maneiras organizacionais que se formam para realização de

pesquisas, que se caracteriza pelo modo de abordagem adotada para cada subcampo específico que em muitos casos, como será possível observar mais a frente neste trabalho, requer uma pesquisa mais abrangente que conseqüentemente necessita de mais pessoas em um grupo ou uma pesquisa mais centrada e específica. Ademais, a coautoria influencia diretamente nas citações, pois se autores de artigos que tem grande destaque, e conseqüentemente receberam alto índice de citações, vão beneficiar com uma parcela das citações totais, autores que colaborarem com o desenvolvimento de artigos em comum.

A análise da relação de colaboração entre os autores neste trabalho visa não somente avaliar a ligação entre grupos de autores, mas também a forma como esses grupos vão se relacionar a partir de tópicos de trabalho, principalmente explorando trabalhos específicos dos autores e suas respectivas citações acumuladas. A visão de cluster mostrada na figura 4 é uma representação ilustrativa que ajuda a entender de maneira simplificada os autores destaque na área de pesquisa abordada, e isso dá a possibilidade de compreensão inicial sobre aspectos importantes que serão tratados nos próximos tópicos deste artigo. Percebeu-se que as redes de colaboração evoluem de acordo com alguns eventos, sendo eles o interesse investigativo, relações institucionais predefinidas e necessidade de conhecimento coletivo. As análises visuais tratadas neste trabalho devem levar em consideração alguns pontos para melhor entendimento: Os nós em destaque são definidos a partir de uma escolha prática da quantidade que deve ser gerada, pensada na visualização simplificada das relações, ou seja, a imagem não contempla todas as possibilidades contidas no texto que, se consideradas, causariam uma grande poluição visual e não compreensão;

destaque se deve pela distribuição de revistas já consolidadas nestes países, que são grandes produtores de artigos pertinentes em diversas áreas, isso causa um grande efeito na busca de repositórios para os trabalhos desenvolvidos, principalmente no que tange o Fator de Impacto da revista que demonstra a sua relevância. A revista PLOS ONE é classificada como a de maior destaque ocupando o primeiro lugar da tabela e acumulou 683 citações para 19 publicações totais até a data de aquisição da base de dados. A revista que ocupa a posição 2 do ranking é denominada BMC Bioinformatics, acumulando 678 citações com 34 publicações. Mas existe uma característica bem peculiar entre essas duas revistas, elas possuem um fator de impacto inferior a todas as outras revistas (3.752 e 3.327 respectivamente), mas como elas possuem um acumulado de citações maior, ocuparam as posições de destaque deste demonstrativo. A revista que ocupa a posição 10, Journal of Molecular Biology (JMB), é a revista que possui um maior quantitativo de média de citações por publicação (95.33 pontos), sendo aproximadamente 300% maior que a PLOS ONE e quase 500% maior que a BMC Bioinformatics, isso revela que esta revista tem potencial para desenvolver mais trabalhos que vão contribuir com a literatura científica. Ao analisar este ranking é possível concluir que as revistas com maiores índices de publicações não necessariamente atraem maior interesse de pesquisa, existe alguns aspectos que podem ser responsáveis por esta classificação, como a facilidade de aceitação no periódico, rapidez de análise e baixo custo. Neste caso específico, probabilisticamente, se as publicações das revistas PLOS One e JMB se equiparassem, haveria um grande destaque de citações da JMB, caso semelhante aconteceria com a ACS Nano.

A segunda seção da tabela 1 é referente aos países que mais publicaram durante o período de análise e seus respectivos dados. Nessa classificação o país que mais obteve citações foi os USA, acumulando 7212 distribuídas nos seus 415 artigos publicados. E a China, ocupando a segunda posição obteve 3903 citações acumuladas nos seus 265 artigos. Uma característica interessante a se acentuar é o fato de que a Inglaterra, que apareceu fortemente

nas revistas científicas, nesta seção ocupa a posição 4. Assim, é possível compreender como se dá essa modulação da diferenciação de análise para cada caso, que pode dar destaque para um país a depender de como ele se desenvolve dentro das estatísticas demonstradas, que pode ser, neste caso, mostrada através da atratividade que as revistas de um país têm para receber solicitações de publicação de autores de diferentes países, contribuindo para o desenvolvimento. Mas, não necessariamente o país da revista vai estar com os mesmos índices classificatórios, isso porque existe a diferenciação entre o país da revista e os países declarados nos artigos, que têm possibilidade de variar pois pode haver autores de diferentes nacionalidades.

A terceira seção da tabela 1 mostra as organizações/instituições que são registradas pelos autores no artigo, podendo haver a possibilidade de serem registrados mais de uma no mesmo trabalho devido a relação de co-autoria. Pode-se observar que a Chinese Academy of Sciences (Chinese Acad Sci) é a que possui a maior quantidade de publicações e citações (36 e 1085 respectivamente) e isso demonstra mais uma vez a força da China nesta área de estudo. Mas, considerando a análise de média de citações por publicação é possível ver que a California Institute of Technology (CalTech) é a que possui maior destaque na tabela, contando com 79.90 citações por artigo, que representa mais que o dobro da primeira colocada citada anteriormente. E por fim, a seção 4 da tabela revela os autores de maior destaque: Wei que publicou 4 artigos e obteve 479 citações (com 119.75 citações por trabalho); Chen que publicou 3 trabalhos e obteve 399 citações (com 133.00 citações por trabalho) e Wang que publicou 3 artigos e acumulou 399 citações (com 133.00 citações por trabalho). Um aspecto importante sobre esta seção é referente ao autor Akutsu que publicou 6 artigos e acumulou apenas cerca de 362 citações, ou seja, mesmo com o dobro de artigos publicados não obteve número de citações iguais ao segundo e terceiro colocado do ranking. Isso mostra a diferenciação de relevância e qualidade dos trabalhos desenvolvidos e sua conformidade com o que é pesquisado dentro da

literatura e se ele se enquadra nos critérios de busca dos pesquisadores.

Tabela 1

Classificação das 10 principais revistas, países, afiliações e autores com base no número de publicações e citações no domínio da Inteligência Artificial em estudos com enzimas.

Rank	Parâmetros	Números				
		<i>Revista</i>	C	FI	NP	NC
1	Plos One	EUA	3.752	19	683	36.47
2	BMC Bioinformatics	ING	3.327	34	678	19.94
3	Journal of Chemical Information Modeling	EUA	6.162	24	633	26.38
4	Proceedings of the National Academy of Sciences of The United States Of America	EUA	12.779	15	571	38.07
5	Bioinformatics	ING	6.931	14	452	32.29
6	Briefings in Bioinformatics	ING	13.994	14	435	31.07
7	ACS Synthetic Biology	EUA	5.249	11	345	31.36
8	Scientific Reports	ING	4.997	25	336	13.44
9	ACS Nano	EUA	18.027	6	306	51.00
10	Journal of Molecular Biology	USA	6.151	3	286	95.33
	<i>País</i>	NP	NC	MC	Total link strength	Porcentagem (%)
1	United States of America	415	7212	17.42	760	32.04%
2	People are China	265	3903	14.73	479	20.51%
3	Germany	104	1769	17.01	385	8.05%
4	England	85	1581	18.60	255	6.58%
5	India	123	1027	8.35	248	9.52%
6	Japan	51	852	16.71	176	3.95%
7	Brazil	45	755	16.78	62	3.48%
8	France	49	744	15.18	203	3.79%
9	Australia	42	718	17.10	132	3.25%
10	Denmark	24	663	27.63	104	1.86%

	<i>Afiliações</i>	NP	NC	MC	Total link strength	Porcentagem (%)
1	chinese acad sci	36	1085	30.13	232	26.47%
2	michigan state univ	10	538	53.80	28	7.35%
3	washington univ	16	451	28.18	107	11.76%
4	kyoto univ	13	450	34.61	123	9.55%
5	monash univ	9	448	49.78	107	6.61%
6	harvard med sch	17	441	25.94	11	12.50%
7	univ cambridge	12	408	34.00	80	8.82%
8	caltech	5	399	79.90	151	3.68%
9	univ sao paulo	9	383	42.55	56	6.61%
10	univ chinese acad sci	9	363	40.33	21	6.61%
	<i>Autores</i>	NP	NC	MC	Total link strength	Porcentagem (%)
1	wei, guo-wei	4	479	119.75	18	9.09%
2	chen, jiahui	3	399	133.00	12	6.81%
3	wang, rui	3	399	133.00	12	6.81%
4	song, jiangning	7	378	54.00	103	15.91%
5	arnold, frances h.	4	373	93.25	214	9.09%
6	akutsu, tatsuya	6	362	60.33	95	13.64%
7	reetz, manfred t.	4	300	75.00	336	9.09%
8	kirchmair, johannes	5	274	54.80	143	11.36%
9	tyzack, jonathan d.	3	269	89.67	99	6.81%
10	swamidass, s. joshua	5	267	53.40	88	11.36%

P = País; FI = Fator de Impacto em 2021; NP = Número de Publicações; NC = Número de

Citações; MC = Média de Citações (NC/NP);. EUA = Estados Unidos da América; ING = Inglaterra; NP = Número de Publicações.

3.1.3 A distribuição por país e instituição

Para melhor compreensão da pesquisa em estudo é necessário entender como geograficamente estão distribuídas as publicações. A figura 5 subdivide-se em 5 partes que vão contribuir com esse conhecimento adicional dos aspectos de pesquisas mencionados. A parte (a) da figura mostra rede colaborativa dos países gerada a partir do programa VOSviewer. Essa parte tem o intuito de ilustrar como ocorre a relação de trabalho em conjunto dos países através dos links que o ligam e, além disso, destacar os países com nível de importância elevado. EUA e China são os de maior destaque e compõem os mesmos grupos colaborativos (chamado de cluster e definido por meio das diferentes cores), isso mostra que essas nações costumam realizar trabalho em conjunto e que obtêm grande respaldo científico. Essa ilustração não cobre todos os países da base de dados devido a quantidade que pode acabar dificultando a visualização, por isso foi selecionado apenas os 35 com maior destaque. A parte (b) se utiliza da mesma base estrutural da parte (a), mas com a diferença conceitual. Neste caso, a parte (b) revela essa relação de colaboração definida através de cores que delimitam um período de análise que mostra especificamente o período de maior concentração de publicações relevantes, que neste caso é representada pelo aumento de citações em ascensão (de 2019 a junho de 2020). Países como os EUA, China, Índia e Alemanha tiveram um pico de citações no início de 2020, já Brasil, Japão, França e Inglaterra obtiveram esse mesmo pico no início de 2019. As causas desses picos podem variar a partir de diversos aspectos como por exemplo a situação política, social e/ou cultural do país em um determinado momento, que vai guiar os rumos e interesses de pesquisa. Uma causa específica que foi citada é a necessidade de investigações científicas devido a pandemia do COVID-19, sendo interesse de pesquisa dos EUA e da CHINA que se destacam na cor verde (referente ao final de 2019 até início de 2020). A parte (c) da figura 5 mostra a distribuição dos artigos por país a partir de uma visualização do mapa global, isso

ajuda a entender melhor que os locais com maior número de pessoas e maior território obtêm um grande destaque nas pesquisas. A figura (d) demonstra a roda de dependência dos países, e isso revela principalmente o nível de relação que cada nação tem com outras, em destaque os EUA que possui forte relacionamento. Essa roda de dependência é desenvolvida a partir de dados coletados no VOSviewer, em especial o campo Total Link Strength (TLS), que é uma representação matemática de similaridade do processamento realizado. Essa similaridade é construída a partir de ocorrências de palavras em documentos, por exemplo se o nome de um autor X ocorrer duas vezes em um determinado artigo com o autor Y significa que entre esses autores existe um TLS igual a 2. Se esse mesmo autor X ocorrer apenas uma vez com um autor Z, significa que o TLS entre esses autores é 1. A roda de dependência se utiliza desses dados para avaliar as relações e assim definir a força de relação entre eles. Por fim, a parte (e) que mostra o gráfico que quantifica as citações dos países mais importantes. Essa visualização complementa a análise feita nas partes a e b da figura, demonstrando com maior precisão principalmente o interesse de pesquisa nos artigos desses países.

ponderadas com outros pontos através de sua força de ligação. Esse modelo de gráfico é muito utilizado para entender o fluxo de dados e pode ser uma excelente forma de ilustrar relações entre dados. e) Gráfico que mostra o nível de citação dos países.

As informações sobre as organizações estão bem ilustradas na figura 6, que se divide em 2 partes. A partir da parte (a) é possível observar as relações de colaboração entre essas instituições, e a partir dessa visualização nota-se que a de maior destaque é a Chinese Academy of Sciences, mesmo com o grande número de publicações e citações dos EUA, o primeiro lugar do pódio desta abordagem vai para uma universidade chinesa. E para uma analogia complementar a parte (b) demonstra a geocodificação das instituições, isso permite identificar a concentração geográfica. Nesta parte os pontos no mapa se acumulam mais no hemisfério norte no globo, principalmente na América do Norte, Europa e Ásia Oriental.

A figura 7 mostra a relação entre autores, países e palavras-chave para assim ser possível entender como acontece a distribuição temática trabalhada por cada autor e de quais países esses autores se originam, que indica o fluxo de trabalhos que estão sendo desenvolvidos por campo, subcampo ou rede de áreas de estudo. A ligação do termo “machine learning” tem ligações diretas com todos os países e conseqüentemente com todos os autores. Algo semelhante acontece com o termo “deep learning”, mas apesar de ter relacionamento com todos os países, há um fluxo de desenvolvimento de trabalhos cerca de 3 vezes menor (indicado pela barra lateral ao nome). O terceiro termo com maior incidência é “covid-19” e isso é um forte indicador do tipo de pesquisa mais pertinente dentro desse campo de estudos das enzimas utilizando IA, principalmente pela pandemia mundial do vírus SARS-CoV-2.

fortemente o acúmulo de coautorias, e as cores divididas em clusters ilustram grupos de instituições com alto grau de colaboração. b) Geocodificação das organizações registradas nos 1292 artigos analisados.

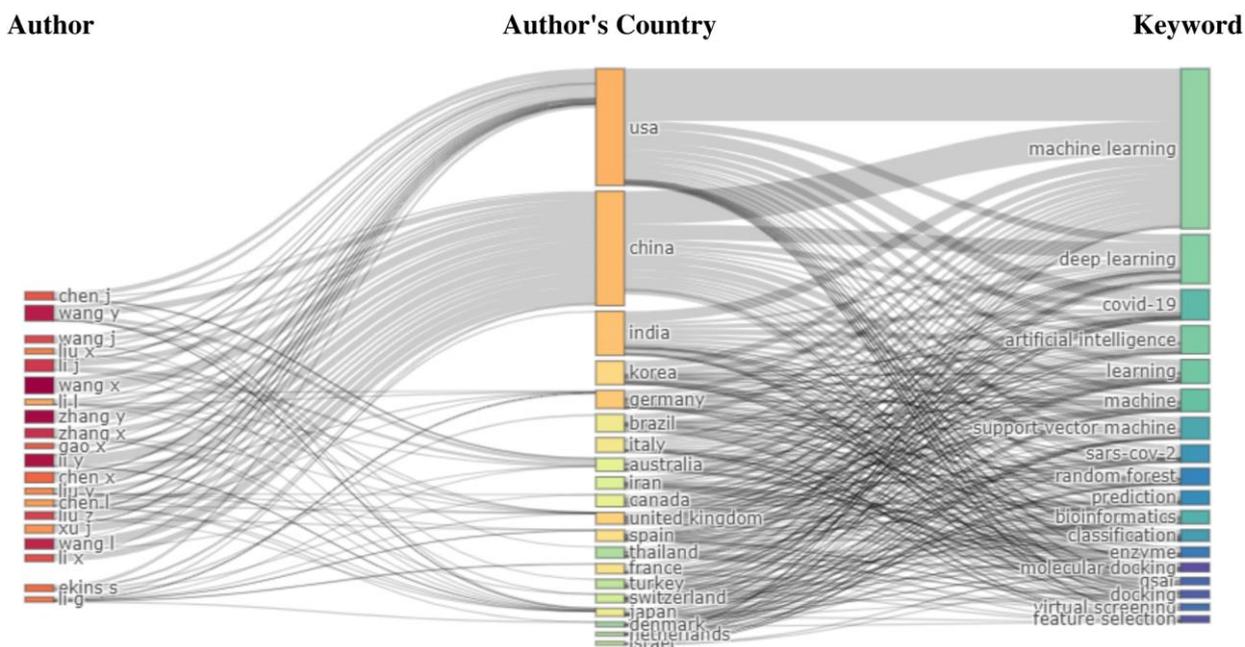


Figura 7: Relação de três campos da pesquisa em estudo (Autor, País do autor e Tópico de pesquisa).

3.1.4 Publicações mais citadas e áreas de investigação

Os artigos mais citados estão listados na tabela 2 junto com seus respectivos autores, anos de publicação e quantidade de citações acumuladas. Analisar essa classificação contribui com o entendimento da variabilidade temática as quais se encaixam no estudo proposto. O primeiro artigo tem como título “Mutations Strengthened SARS-CoV-2 Infectivity” (Chen *et al.*, 2020) desenvolvido por Chen, este artigo tem como objetivo aplicar o TopNetTree (técnica de ML) para analisar os impactos das mutações no domínio de ligação ao receptor (RBD) da proteína S existentes na energia livre de ligação (BFE) da proteína S e do angiotensina 2 (ACE2). A análise deste artigo ajuda a compreender melhor que uma abordagem específica do

estudo com enzimas pode-se utilizar do auxílio das tecnologias de ML. O segundo artigo de título “Cellphone-Based Hand-Held Microplate Reader for Point-of-Care Testing of Enzyme-Linked Immunosorbent Assays” (Berg *et al.*, 2015) do autor Berg vai falar sobre o desenvolvimento de um leitor colorimétrico baseado em smartphone de baixo custo e portátil utilizado no monitoramento rápido de imunoenaios imunossorventes enzimáticos (ELISA) colorimétricos baseados em placas de microtitulação (MTP). O terceiro artigo intitulado “Toward more realistic drug–target interaction predictions” (Pahikkala *et al.*, 2015) do autor Pahikkala trata do uso de modelo de machine learning para previsão de interações droga-alvo baseados na estrutura química e nas informações da sequência genômica. O quarto artigo com título “Computational Prediction of Metabolism: Sites, Products, SAR, P450 Enzyme Dynamics, and Mechanisms” (Kirchmair *et al.*, 2012) do autor Kirchmair vai falar sobre os diversos métodos computacionais aplicados à previsão do metabolismo e também realizar o processo comparativo e de contraste para a sua aplicabilidade e desempenho. O quinto artigo tem como título “Towards an Artificial Intelligence Framework for Data-Driven Prediction of Coronavirus Clinical Severity” (Jiang *et al.*, 2020) do autor Jiang, que trata da construção de uma estrutura de IA utilizando recursos de análise preditiva aplicados a dados reais de pacientes, para fornecer suporte rápido à tomada de decisões clínicas.

O sexto artigo denominado “Machine learning-assisted directed protein evolution with combinatorial libraries”(Wu *et al.*, 2019) do autor Wu vai mostrar a incorporação do aprendizado de máquina ao fluxo de trabalho de evolução direcionada para tentar mitigar esforço experimental relacionado à evolução de proteínas direcionadas e examinar o espaço de sequência codificado pela mutação de diversas posições paralelamente. O sétimo artigo com título “Recent advances in electronic tongues”(Riul Jr. *et al.*, 2010) do autor Riul, descreve os desenvolvimentos destaque das línguas eletrônicas (e-línguas) e dos sensores gustativos recentemente, com uma amostragem dos métodos de detecção e materiais usados nas unidades

de detecção utilizando IA para melhoramento do desempenho. O oitavo artigo, “The Crucial Role of Methodology Development in Directed Evolution of Selective Enzymes”(Qu *et al.*, 2020) desenvolvido por Qu, se concentra em explicar os desenvolvimentos recentes sobre otimização de métodos de biologia molecular para mutagênese gênica e o desenho de estratégias eficientes no campo da evolução dirigida de enzimas estéreo, regio e quimiosseletivas. O nono artigo de Baloglu, denominado “Classification of myocardial infarction with multi-lead ECG signals and deep CNN”(Baloglu *et al.*, 2019) vai tratar de um modelo de aprendizado profundo composto por uma estrutura de ponta a ponta no sinal de eletrocardiograma padrão de 12 derivações para o diagnóstico de infarto do miocárdio, de maneira mais específica usando a rede neural convolucional. E por fim o décimo artigo da tabela que tem como título “Predicting Drug-Target Interactions for New Drug Compounds Using a Weighted Nearest Neighbor Profile” (van Laarhoven; Marchiori, 2013) do autor van Laarhoven que discute um método que prever pares de interação com alta precisão para esses compostos de drogas (ou alvos) não catalogados. O trabalho demonstra um procedimento de vizinho mais próximo ponderado simples que é altamente eficaz para esta atividade. Ao avaliar este ranking compreende-se que existe algumas características que devem ser observadas, como a relação entre ano de publicação e quantidade de citação, pois a partir disso pode-se realizar uma associação da possibilidade de qualidade e interesse do artigo. O artigo que ocupa a primeira posição que foi publicado em 2020 e já obteve 269 citações, isso pode identificar aspectos atrativos como qualidade, exclusividade de métodos ou estudos, inovação ou relevância. Como este artigo trata sobre o SARS-CoV-2, com o aumento do interesse de pesquisa devido a pandemia desse vírus iniciada em 2019, entende-se os números excelentes. Com o destaque desses artigos compreende-se que eles servirão como referência para muitos estudos, e assim contribuindo mais na modelagem do conhecimento científico.

E diante do que foi discutido, os artigos se subdividem em vários campos específicos

da ciência, que é ilustrada na figura 8. Essa figura mostra que a área de estudo com maior quantidade de artigos publicados é a “Biochemistry Molecular Biology” com 374 registros e representando cerca de 28.9% dos registros de todas as áreas juntos. logo em seguida a “Chemistry” com um acumulado de 271 registros, representando aproximadamente 20.9% do montante, e em terceiro lugar “Computer Science” com um total de 194 registros e com cerca de 15% do valor total. Essas áreas representam a rede de bases científicas de pesquisa, sendo evidenciada pela necessidade individual de cada autor ou grupo. Cada campo de pesquisa ilustrado na tem uma gama de interesses investigativos para as várias abordagens de IA nos estudos com enzimas. Ao realizar a mesma pesquisa na WoSCC e refinando apenas para a área “Biochemistry Molecular Biology” é possível obter os artigos pertencentes a ela. Por exemplo, além do aparecimento do artigo da posição 1 da tabela 2, encontra-se em destaque o artigo de Miao Mei, “Current Strategies of Antiviral Drug Discovery for COVID-19” (Mei; Tan, 2021), que fala sobre a descoberta e o desenvolvimento de novas drogas para a infecção por SARS-CoV-2 com base nas estratégias das quais derivam os novos medicamentos, sendo uma delas as metodologias assistidas por ferramentas de IA. Ao correlacionar os dois artigos, percebe-se o grande interesse na área sobre a pandemia do COVID-19 e descoberta de medicamentos, que não se restringe a apenas isto. Este fato pode ser revelado observando mais amostras, como o artigo intitulado “Biosystems Design by Machine Learning” (Volk *et al.*, 2020) do autor Michael Jeffrey Y. Volk, que discute a utilização de um modelo de ML para o design de biosistemas (enzimas, vias e células inteiras), ou seja, um outro paradigma extremamente diferente. Mas para ajudar a diferenciar pode-se observar os artigos da área “Computer Science”, escolhido propositalmente devido a aparente diferença científica entre a duas. O artigo classificado pela plataforma como mais relevante é chamado “ProPythia: A Python package for protein classification based on machine and deep learning” (Sequeira; Lousa; Rocha, 2022) da autora Ana Marta Sequeira, neste trabalho é proposto o ProPythia, um pacote

Python genérico que permite implantar de maneira facilitada métodos de ML e DL para uma variabilidade de problemas na análise e classificação de sequências de proteínas.

Tabela 2

Os artigos mais citados na área de investigação.

Rank	Artigo	Autores	Ano de Publicação	Citação
1	Mutations Strengthened SARS-CoV-2 Infectivity(Chen <i>et al.</i> , 2020)	Chen, Jiahui; Wang, Rui; Wang, Menglun; Wei, Guo-Wei	2020	269
2	Cellphone-Based Hand-Held Microplate Reader for Point-of-Care Testing of Enzyme-Linked Immunosorbent Assays(Berg <i>et al.</i> , 2015)	Berg, Brandon; Cortazar, Bingen; Tseng, Derek; Ozkan, Haydar; Feng, Steve; Wei, Qingshan; Chan, Raymond Yan-Lok; Burbano, Jordi; Farooqui, Qamar; Lewinski, Michael; Di Carlo, Dino; Garner, Omai B.; Ozcan, Aydogan	2015	250
3	Toward more realistic drug-target interaction predictions(Pahikkala <i>et al.</i> , 2015)	Pahikkala, Tapio; Airola, Antti; Pietila, Sami; Shakyawar, Sushil; Szwajda, Agnieszka; Tang, Jing; Aittokallio, Tero	2015	217
4	Computational Prediction of Metabolism: Sites, Products, SAR, P450 Enzyme Dynamics, and Mechanisms(Kirchmair <i>et al.</i> , 2012)	Kirchmair, Johannes; Williamson, Mark J.; Tyzack, Jonathan D.; Tan, Lu; Bond, Peter J.; Bender, Andreas; Glen, Robert C.	2012	202
5	Towards an Artificial Intelligence Framework for Data-Driven Prediction of Coronavirus Clinical Severity(Jiang <i>et al.</i> , 2020)	Jiang, Xiangao; Coffee, Megan; Bari, Anasse; Wang, Junzhang; Jiang, Xinyue; Huang, Jianping; Shi, Jichan; Dai, Jianyi; Cai, Jing; Zhang, Tianxiao; Wu, Zhengxing; He, Guiqing; Huang, Yitong	2020	199
6	Machine learning-assisted directed protein evolution with combinatorial libraries(Wu <i>et al.</i> , 2019)	Wu, Zachary; Kan, S. B. Jennifer; Lewis, Russell D.; Wittmann, Bruce J.; Arnold, Frances H.	2019	188
7	Recent advances in electronic tongues(Riul Jr. <i>et al.</i> , 2010)	Riul, Antonio, Jr.; Dantas, Cleber A. R.; Miyazaki, Celina M.; Oliveira, Osvaldo N., Jr.	2010	186
8	The Crucial Role of Methodology Development in Directed Evolution of Selective Enzymes(Qu <i>et al.</i> , 2020)	Qu, Ge; Li, Aitao; Acevedo-Rocha, Carlos G.; Sun, Zhoutong; Reetz, Manfred T.	2020	184

9	Classification of myocardial infarction with multi-lead ECG signals and deep CNN(Baloglu <i>et al.</i> , 2019)	Baloglu, Ulas Baran; Talo, Muhammed; Yildirim, Ozal; Tan, Ru San; Acharya, U. Rajendra	2019	181
10	Predicting Drug-Target Interactions for New Drug Compounds Using a Weighted Nearest Neighbor Profile(van Laarhoven; Marchiori, 2013)	van Laarhoven, Twan; Marchiori, Elena	2013	163

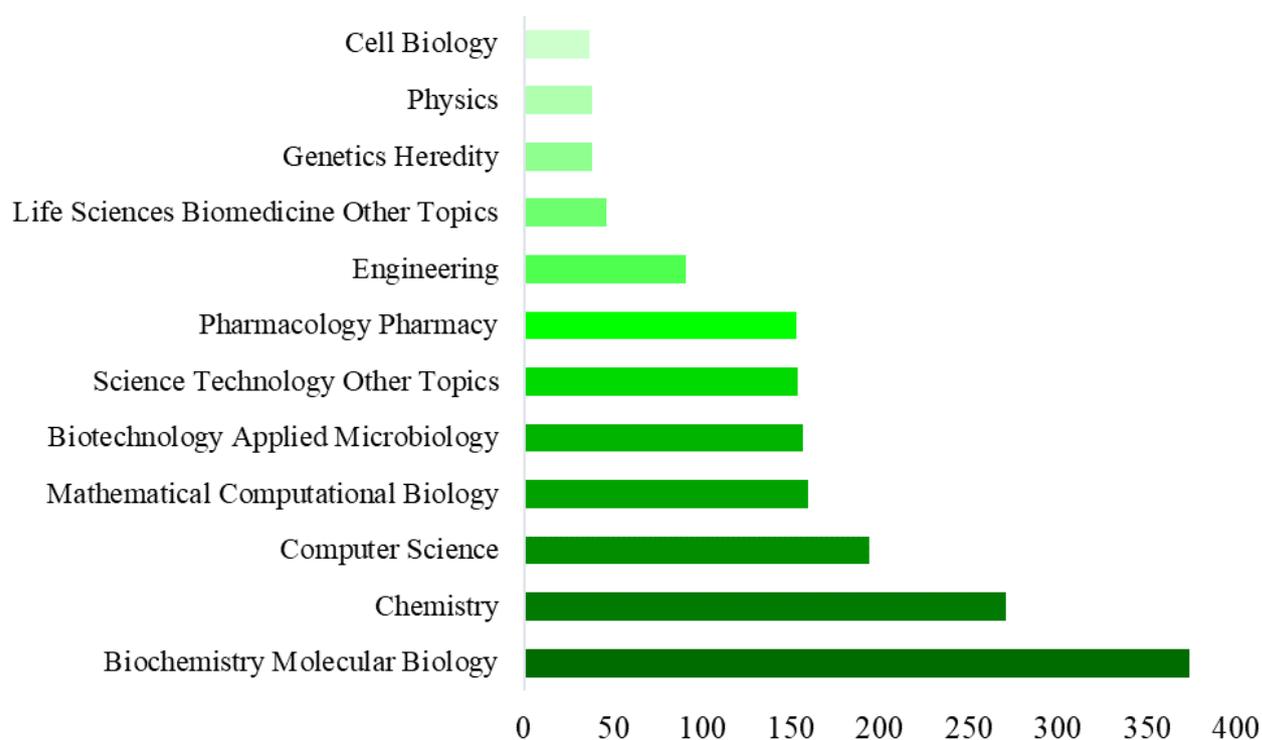


Figura 8: Distribuição das áreas de pesquisa definidas na base de dados em investigação.

3.2 Overview das principais utilizações da IA nos processos de engenharia enzimática

A engenharia enzimática é o processo de melhorar a eficiência de uma enzima já disponível ou a formulação de uma atividade enzimática avançada, alterando sua sequência de aminoácido (Basheer; Chellappan, 2017; Vanella *et al.*, 2022). A figura 9 mostra como uma abordagem genérica de machine learning acontece no processamento de uma enzima no meio computacional. É possível observar que as técnicas experimentais existentes e as técnicas

computacionais podem não ser capazes de amostrar todo o espaço da sequência de proteínas e se beneficiar completamente do potencial da natureza para a geração de melhores enzimas (Siedhoff; Schwaneberg; Davari, 2020). Com os avanços no sequenciamento de próxima geração, métodos de triagem de alto rendimento, o crescimento de bancos de dados de proteínas e inteligência artificial, especialmente aprendizado de máquina (ML), a engenharia de enzimas orientada por dados está surgindo como uma solução promissora para esses desafios (Siedhoff; Schwaneberg; Davari, 2020; Steiner; Schwab, 2012). Vários algoritmos de ML já foram aplicados à engenharia de enzimas, como por exemplo: florestas aleatórias aplicadas para prever a solubilidade de proteínas, máquinas vetoriais de suporte e árvores de decisão para prever mudanças na estabilidade da enzima após mutações, classificadores K-vizinho Mais Próximo (KNN) para prever a função da enzima e mecanismos, e vários algoritmos de pontuação e agrupamento para anotação rápida de sequência funcional (Jiang; Ran; Yang, 2023). A principal atratividade do ML na engenharia de enzimas decorre de sua generalização: uma vez treinado na entrada conhecida, chamada de conjunto de treinamento, um algoritmo de ML pode potencialmente fazer previsões sobre novas variantes quase instantaneamente (Derat; Kamerlin, 2022; Mazurenko; Prokop; Damborsky, 2020). a ML foi utilizada para responder a questões antigas em enzimologia: realizar a predição da função (ou funções) de uma enzima putativa a partir de sua sequência ou estrutura, prever de que maneira as mudanças ambientais afetarão a enzima função e prever quais mutações melhorariam as taxas de renovação catalítica (Fehan; Montezano; SG Slusky, 2021).

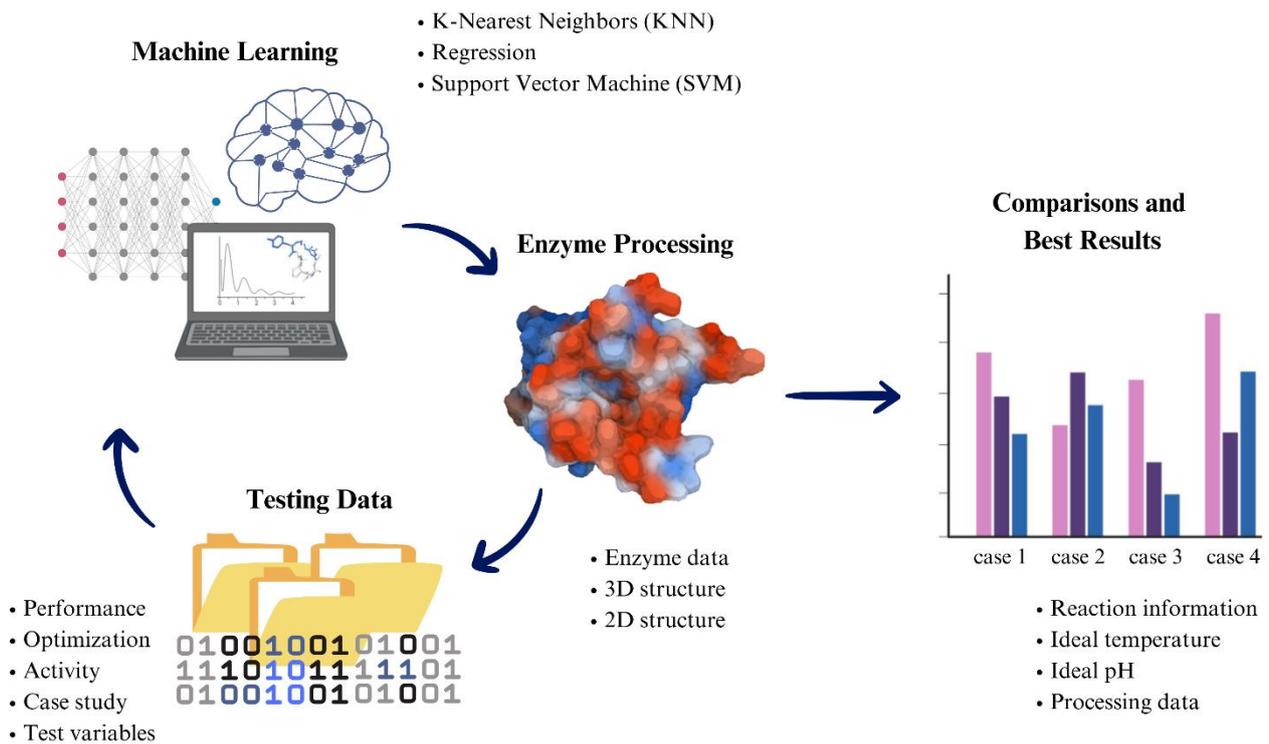


Figura 9: Funcionamento genérico de um algoritmo de machine learning no tratamento computacional de uma enzima.

3.2.1 Previsão de enzimas utilizando Machine Learning

Diante das informações discutidas até aqui um tópico de extrema importância vista em diversos artigos foi a utilização de técnicas e ferramentas de ML para realizar predição em alguns processos utilizando enzimas, como o trabalho de Jiang (Jiang *et al.*, 2020) e o de van Laarhoven (van Laarhoven; Marchiori, 2013). Um campo muito importante é a predição da interação droga-enzima que, primordialmente, é um procedimento demorado e caro. É muito difícil de se conseguir encontrar novas enzimas-alvo e novas drogas candidatas somente por meio de experimentos bioquímicos tradicionais (Niu *et al.*, 2014). Além disso, existe o processo de predição da localização de locais ativos em enzimas (Gutteridge; Bartlett; Thornton, 2003). A previsão de sítios ativos de enzimas em proteínas é uma questão de extrema relevância não somente para as ciências de proteínas, mas também para uma diversidade de aplicações práticas, como por exemplo o design de drogas. Como os mecanismos de reação enzimática são baseados

nas estruturas locais dos sítios ativos da enzima, vários métodos baseados em modelos que comparam estruturas locais em proteínas foram desenvolvidos até o momento (Kato; Nagano, 2011).

Outro uso do ML neste campo é prever a função da proteína a partir da estrutura. Estes métodos de predição da função da proteína a partir da estrutura estão se tornando mais importantes à medida que a taxa na qual as estruturas são resolvidas evolui de maneira mais acelerada do que o conhecimento experimental (Dobson; Doig, 2005). A tabela 3 vai demonstrar alguns métodos de aprendizado de máquina que são usados como suporte para construção de soluções e processos que se envolvem com o campo das enzimas. Os itens da tabela 3 são aplicados a diversas metodologias que podem ou não ser utilizadas em conjuntos, tendo em vista a necessidade específica de cada abordagem como é referenciada nos artigos da terceira coluna.

Tabela 3

Métodos de aprendizado de máquina no campo da predição das abordagens enzimáticas.

Método	Aplicação	Referência
Artificial Neural Networks (ANN)	Prediction of enzyme class, activity, specificity and stability; modeling of enzymatic reactions and metabolic pathways	(Bitam; Hamadache; Hanini, 2023; Concu <i>et al.</i> , 2022; de Menezes <i>et al.</i> , 2023; Ghose <i>et al.</i> , 2023; Kumar, Gaurav <i>et al.</i> , 2022; Levin <i>et al.</i> , 2022; Ming <i>et al.</i> , 2023a; Upadhyay; Boorla; Maranas, 2023)
Support Vector Machines (SVM)	Prediction of enzyme class, function, family and active site; identification of enzymes in genomic sequences	((Hogan; Peut, 2014; Huang; Lu; Liu, 2023; Kroll <i>et al.</i> , 2023; Nourouzi; Chamani, 2021; Zhang, R. <i>et al.</i> , 2023))
Decision Trees (DT)	Prediction of enzyme class, function and family; analysis of gene expression and enzyme metabolism	((Ebrahimi <i>et al.</i> , 2011; Ebrahimi; Ebrahimie, 2010; Gao <i>et al.</i> , 2013; Khan; Hayat; Khan, 2015; Kirchmair <i>et al.</i> , 2013; Syed; Yona, 2009))
Genetic Algorithms (GA)	Optimization of enzyme parameters and structures; design of new enzymes with desired properties	((Dixit; Chhabra; Shukla, 2023; Ousaadi <i>et al.</i> , 2021; Zhang <i>et al.</i> , 2012; Zhang; Lin; Chen, 2023b))

Bayesian Regularization (BR)	Inference of causal relationships between genes, enzymes and metabolites; reconstruction and analysis of metabolic networks	((Fernández; Caballero, 2006; Fernandez; Caballero, 2016; Hosseini; Javidanbardan; Khatami, 2019))
K-Means (KM)	Grouping enzymes based on similar characteristics; identifying patterns and outliers in enzyme data	((Roy; Roy, 2009; Szaleniec, 2012))
Logistic Regression (LR)	Binary classification of enzymes based on numerical or categorical characteristics; assessing the probability of an enzyme belonging to a particular class or function	(Ahuja <i>et al.</i> , 2000; Donati, 2004; Shahid; Masood; Khan, 2021)
Principal Component Analysis (PCA)	Dimensionality reduction and visualization of enzyme data; identification of the most relevant characteristics to explain the variance of the data	(Ajjolli Nagaraja <i>et al.</i> , 2020; Sharif <i>et al.</i> , 2016; Xu; Van Doren, 2018)
Random Forests (RF)	Prediction of enzyme class, function and family; analysis of the importance of characteristics for prediction; combination of various decision tree models	(Kumar; Choudhary, 2012; Kumar; Li; Choudhary, 2009; Mou <i>et al.</i> , 2021)
Deep Learning (DL)	Prediction of the three-dimensional structure, function and interaction of enzymes; recognition of complex patterns in high-dimensional data; use of multiple layers of artificial neural networks	(Chehili <i>et al.</i> , 2021; Duhan; Norton; Kaundal, 2022; Ryu; Kim; Lee, 2019)

A ANN pode ser visualizada em diversos trabalhos e de várias formas e abordagens. O artigo “A novel enzyme based biosensor for catechol detection in water samples using artificial neural network” (Maleki *et al.*, 2017) do autor Nasim Maleki utiliza esta técnica para desenvolver um novo sistema integrado de biossensor/ANN usando Poli (3,4-etilenodioxifeno) (PEDOT), nanofolhas de óxido de grafeno (GONs) e lacase (Lac) no desenvolvimento do biossensor. É falado que a ANN tem a capacidade de modelar sinais lineares ou não lineares com alta precisão, operações adaptativas em tempo real e calcular o resultado em um curto período. No artigo de Debasmita Pradhan, “Enzyme classification using multiclass support vector machine and feature subset selection” (Pradhan; Padhy; Sahoo, 2017), um modelo composto por seleção de subconjunto de recursos seguido por SVM é proposto para determinar a classe funcional de uma sequência de proteína. Aqui são consideradas 32 propriedades físico-químicas de enzimas de 6 classes de enzimas para realização de treinamento da ferramenta. No

artigo “Cost-effectiveness Assessment of Commonly-used Drugs for Hepatoprotection and Enzymes Reduction Based on Decision Tree and Multi-utility Theory” (Xu *et al.*, 2014) do autor Jia-qi Xu, foi usado a DT para analisar dados da relação custo-efetividade dos três grupos de 70 casos de pacientes que foram submetidos a 3 protocolos terapêuticos (composto glicosídeo glicirrízico combinado com injeção de ornitina aspártica, composto glicosídeo glicirrízico combinado com fosfatidilcolina e composto glicosídeo glicirrízico combinado com tiopronina). É falado que DT é um modelo desenvolvido para listar todos os resultados potenciais das escolhas, com o objetivo de fornecer estatísticas base para a tomada de decisão. O artigo de M. Mohebbi, “Modeling and optimization of viscosity in enzyme-modified cheese by fuzzy logic and genetic algorithm” (Mohebbi *et al.*, 2008), utiliza o GA como ferramenta complementar na construção de um Sistema Genético Fuzzy de Base de Regras (GFRS) para modelagem de viscosidade em queijo modificado por enzima (EMC). No documento “Genetic Algorithm Optimization of Bayesian-Regularized Artificial Neural Networks in Drug Design” (Fernandez; Caballero, 2016) de Michael Fernandez é aplicado redes neurais artificiais reguladas bayesianas (BRANNs) combinadas com seleção de recursos de algoritmo genético (GA), nomeadas redes neurais genéticas reguladas bayesianas (BRGNNs), em relações quantitativas a implementação de redes neurais artificiais reguladas bayesianas (BRANNs) combinadas com seleção de recursos de algoritmo genético (GA), as chamadas redes neurais genéticas reguladas bayesianas (BRGNNs), em análise de relações quantitativas/qualitativas estrutura-atividade (QSAR). O BRGNN foi incorporado ao design auxiliado por computador de candidatos a medicamentos para uma variedade de doenças (doenças cardíacas, AIDS, câncer, infecções fúngicas, etc.). Neste trabalho, as redes neurais foram treinadas para calcular as atividades biológicas de um amplo espectro de medicamentos.

No artigo “The Influence of Enzymatic Hydrolysis on Bee Pollen Antioxidant and Antibacterial Activities”(Damulienė *et al.*, 2023) de Vaida Damulienė, visa determinar o

impacto da hidrólise enzimática nas atividades antioxidantes e antibacterianas do pólen de abelha, utilizando a análise de agrupamento K-means para clusterizar amostras em 5 a 9 grupos. No artigo de Eva Csoz, “Substrate Preference of Transglutaminase 2 Revealed by Logistic Regression Analysis and Intrinsic Disorder Examination” (Csoz *et al.*, 2008), é demonstrado um estudo sobre a preferência do substrato da enzima transglutaminase tecidual (TG2) no nível da estrutura terciária, são comparados dados derivados do ambiente tridimensional (3D) de resíduos de substrato e não-substrato de proteínas de substrato cristalizadas. Com a aplicação da análise de LR, foi possível identificar resíduos de aminoácidos preditores que são importantes na discriminação entre resíduos substrato e não substrato. Ainda, no artigo é descrito que a LR é um método estatístico originalmente utilizado para prever a probabilidade de ocorrência de um evento com a ajuda de diferentes preditores.

O autor Jia Xu no artigo “Affinities and Comparisons of Enzyme States by Principal Component Analysis of NMR Spectra, Automated using TREND Software” (Xu; Van Doren, 2018), apresenta aplicações potenciais do software TREND NMR para amostras únicas que sofrem alterações, dando destaque para aplicações relação de ligantes com comportamentos desafiadores para Ressonância Magnética Nuclear (RMN). Realiza-se comparações e agrupamentos de estados de enzimas detectados por RMN, usando PCA. No artigo “Comparison of random forest and multiple linear regression models for estimation of soil extracellular enzyme activities in agricultural reclaimed coastal saline land”(Xie *et al.*, 2021) do autor Xuefeng Xie é tratado da utilização modelos de regressão linear múltipla (MLR) e floresta aleatória (RF) para estimar as atividades das enzimas amilase e urease do solo usando covariáveis, como conteúdo de nitrogênio total, água no solo, fósforo total, condutividade elétrica e carbono orgânico do solo, e também características como pH e densidade do solo. No artigo “Enzyme Function Prediction using Deep Learning”(Memon; Aamir Khan; Naveed, 2020) do autor Safyan Aman Memon aplica abordagens inovadoras de aprendizagem profunda

atrelado com um esquema de classificação hierárquica eficiente para prever a função de uma enzima. Tendo em vista todos os exemplos de usos de cada técnica de ML, é possível entender a aplicabilidade real no mundo da engenharia enzimática.

3.2.2. Aplicações de Inteligência Artificial no design de enzimas

A inteligência artificial (IA) disponibilizou recentemente soluções para modelagem estrutural, observando pontos críticos de alteração e orientando mutações para necessidades específicas, o que ajuda muito a engenharia enzimática (Singh *et al.*, 2021). As ferramentas baseadas em IA mostraram grandes vantagens na redução do tempo computacional, permitindo maior precisão de previsão e fornecendo modelos treináveis adequados para usos amplos (Wang, Xinglong *et al.*, 2023). Um dos desafios mais fundamentais no design de proteínas com IA é como representar a proteína em formatos legíveis por máquina (Jang *et al.*, 2022b; Wang *et al.*, 2021). Embora uma rede neural profunda possa extrair automaticamente padrões complexos de dados em grande escala sem conhecimento prévio, uma representação bem escolhida pode tornar o aprendizado de máquina mais eficaz (Jang *et al.*, 2022b).

Dentre as aplicações de IA de muita relevância no design de enzimas estão: O projeto da estabilidade catalítica enzimática (Ming *et al.*, 2023a); Previsão da estrutura das proteínas (Chungyoun; Gray, 2023; Roche *et al.*, 2023); Evolução dirigida (Jang *et al.*, 2022b); Design Racional (Paladino *et al.*, 2017); Triagem virtual (Cañizares-Carmenate *et al.*, 2022; Paria; Tassanakajon, 2023); Otimização de processos (Dixit; Chhabra; Shukla, 2023); Seleção de enzimas de alta eficiência (Sampaio; Fernandes, 2023). Nos últimos anos, o design de proteínas auxiliado por IA tornou-se uma abordagem cada vez mais conhecida que possibilita a otimização das funções das proteínas (Jang *et al.*, 2022c).

3.2.3 Otimização de Processos

A otimização de processos geralmente envolve-se a uma abordagem simples e convencional. O processo demonstra grandes desvantagens por ser demorado, impreciso, com

previsões pseudo-ótimas e por não estudar as interações de diferentes variáveis do processo que afetam o rendimento final (Haque *et al.*, 2016). As ferramentas de inteligência artificial, em especial o aprendizado de máquina, é muito utilizada para desenvolver e otimizar problemas práticos de engenharia (Aydın Temel; Cagcag Yolcu; Turan, 2023). Baseados na análise de big data, os algoritmos de aprendizado de máquina têm a vantagem da previsão com alta precisão e da capacidade de resolver problemas não lineares, quando comparado com os métodos de análise tradicionais (Wang, Youzhaoh *et al.*, 2023). Os modelos baseados em dados de IA não requerem os mecanismos dos processos bioquímicos e são mais poderosos para simular processos complexos e intrinsecamente conectados que podem ser usados para fins de previsão operacional (Luo, Jingyang *et al.*, 2023).

Quando selecionado um algoritmo adequadamente preciso para modelar a relação estrutura-função, ele recebe treinamento, para que os parâmetros sejam ajustados para potencializar sua capacidade preditiva, e avaliado para analisar seu desempenho (Sampaio; Fernandes, 2023). Posteriormente, a otimização baseada em modelo pode prosseguir para a seleção da sequência ou sequências que otimizam a função (Feehan, Ryan; Montezano, Daniel ; Slusky, 2021; Sampaio; Fernandes, 2023). O algoritmo genético (AG), um dos algoritmos de otimização metaheurística, tem como base o processo de seleção natural que é possível observar na natureza (Kumi; Jeong, 2023; Mitchell, 1998). É frequentemente utilizado como um algoritmo aleatório para resolver problemas de otimização com condições restritas (Tung *et al.*, 2023). GA simula essencialmente a forma como a vida evolui para encontrar soluções para problemas do mundo real (Mitchell, 1998).

3.2.4 *Descoberta de novas enzimas*

A descoberta de novas enzimas é necessária para expandir os caminhos que podem ser acessados pela engenharia metabólica para a biossíntese de compostos funcionais (Watanabe *et al.*, 2020, 2022). Consequentemente, vários modelos de aprendizado de máquina

foram desenvolvidos para prever reações enzimáticas (Ge *et al.*, 2023; Watanabe *et al.*, 2020). A descoberta de enzimas depende fortemente de uma base de conhecimento bem estabelecida de funções enzimáticas, tornando a previsão da função proteica crucial para expandir a gama de enzimas aplicáveis (Ryu *et al.*, 2023).

Embora o aprendizado de máquina tenha recebido muito entusiasmo recentemente ele não é uma ferramenta milagrosa (Bergman *et al.*, 2023; Manimaran *et al.*, 2024). Uma desvantagem notável é que as técnicas de ML necessitam de muita informação (Dharmik; Bawankar, 2023). Mesmo os modelos de aprendizado de máquina mais sofisticados são tão bons quanto a quantidade e a qualidade dos dados de treinamento subjacentes. Na verdade, o aumento da complexidade do modelo requer maiores quantidades de dados (Robinson; Piel; Sunagawa, 2021). O aprendizado profundo, em particular, comumente sofre de overfitting (ajuste excessivo), ou seja, não podem ser generalizadas para estudos diferentes ou conjuntos de dados (Kadhim; Abdullah; Ghathwan, 2023). Portanto, o apoio contínuo e a curadoria de bancos de dados públicos que fornecem dados de treinamento de alta qualidade, como Swiss-Prot (Bairoch, 2000; Bairoch; Boeckmann, 1994; Karp; Paley; Zhu, 2001) e MIBiG (Kautsar *et al.*, 2019; Terlouw *et al.*, 2023) são essenciais para o aprendizado de máquina, a fim de permitir futuras descobertas de enzimas (Robinson; Piel; Sunagawa, 2021).

3.2.5 Estabilidade das enzimas

As enzimas sofrem uma variedade de reações de desnaturação durante a produção, armazenamento e aplicação na indústria (Iyer; Ananthanarayan, 2008). A desnaturação é o desdobramento da estrutura terciária da enzima em um polipeptídeo desordenado no qual os resíduos principais não estão mais alinhados o suficiente para a participação contínua em interações funcionais ou de estabilização de estrutura. Pode ser revertido se a influência desnaturante for removida (Fágáin, 1995). Modelos matemáticos eficazes que descrevem a estabilidade da enzima sob variadas condições são de extrema importância para prever o

rendimento do biocatalisador (quantidade total de produto produzido para uma determinada carga de biocatalisador) e, portanto, a contribuição do custo do biocatalisador para os custos operacionais gerais (Woodley, 2019). Para garantir a atividade, funcionalidade e estabilidade adequadas da enzima, é essencial compreender sua distribuição espacial e química quando imobilizada (Ralbovsky; Smith, 2021). As caracterizações virtuais de enzimas com o auxílio de algoritmos de aprendizado de máquina oferecem uma grande oportunidade para melhorar as classificações sistemáticas (Damborsky; Brezovsky, 2014; Sraphet; Javadi, 2021). Eles podem ajudar nas propostas científicas de rotina para desenvolver enzimas melhores com atividade superior. Além disso, a análise computacional oferece uma maneira mais clara de compreender os mecanismos de cada reação do ponto de vista estrutural (Sraphet; Javadi, 2022).

O projeto da estabilidade catalítica enzimática é de grande importância na medicina e na indústria. No entanto, os métodos tradicionais são demorados e caros (Ming *et al.*, 2023b). Assim, um número crescente de ferramentas computacionais complementares tem sido desenvolvido, por exemplo: FireProt (Bednar *et al.*, 2015; Liu *et al.*, 2021; Maenpueen *et al.*, 2020), AlphaFold2 (Casadevall *et al.*, 2022; Casadevall; Duran; Osuna, 2023; Jiao; Xing; Sun, 2023; van der Weg; Gohlke, 2023), RoseTTAFold (Golinelli-Pimpaneau, 2021; Watson *et al.*, 2023), ProteinMPNN (Dauparas *et al.*, 2022; Kao *et al.*, 2023) e ESMFold (Meng; Guo; Tang, 2023; Valdés-Tresanco *et al.*, 2023). Eles são propostos para design de enzimas guiado por algoritmos e dados através de metodologias de IA incluindo processamento de linguagem natural, aprendizado de máquina, aprendizado profundo, rede neural de passagem de mensagens e autocodificador variacional/rede adversária generativa (MPNN) (Ming *et al.*, 2023b). Além disso, os desafios do projeto da estabilidade catalítica enzimática englobam dados estruturados insuficientes, grande espaço de busca de sequências, previsão quantitativa imprecisa, baixa eficiência na validação experimental e um processo de projeto complicado (Mazurenko; Prokop; Damborsky, 2020). O primeiro princípio do projeto de estabilidade

catalítica enzimática é tratar os aminoácidos como o elemento básico. Ao projetar a sequência de uma enzima, a flexibilidade e a estabilidade da estrutura são ajustadas, controlando assim a estabilidade catalítica da enzima num ambiente industrial específico ou num organismo (Li *et al.*, 2021). Indicadores comuns dos objetivos do projeto incluem a mudança na energia de desnaturação (Ahmad, 2022), temperatura de fusão (Korkegian *et al.*, 2005), temperatura ideal (Gomes *et al.*, 2007), pH ideal (pHopt) (Pallavi *et al.*, 2010), etc .

As estratégias de estabilização abrangem uma ampla gama de métodos baseados no empirismo (SCHYMKOWITZ *et al.*, 2005) ou funções de energia física (Alford *et al.*, 2017), utilização da estatística (Dehouck *et al.*, 2009), cálculo da energia livre de proteínas (Wickstrom; Gallicchio; Levy, 2012), abordagens de bioinformática baseadas no design de consenso (Porebski; Buckle, 2016) ou reconstrução de sequência ancestral (Musil *et al.*, 2021), e algoritmos contemporâneos de ML (Folkman *et al.*, 2016; Shroff *et al.*, 2020). Cada abordagem tem os seus méritos e também os seus pontos fracos, e a avaliação e escolha da mais adequada não é uma tarefa trivial que exige uma experiência considerável para a sua aplicação bem sucedida (Kunka *et al.*, 2023).

3.2.6 Automação de laboratórios

A automação do laboratório envolve relacionar os conceitos abstratos dos procedimentos químicos e o hardware que ficará responsável pela execução (Hammer *et al.*, 2021; Wilbraham; Mehr; Cronin, 2021). Isto pode ser obtido por meio da criação de uma representação virtual completamente conectada do equipamento físico e do seu estado, ou seja, uma “gêmeo digital” do laboratório que preenche a lacuna entre o mundo virtual e o mundo real. o Gêmeo Digital (DT) foi proposto para promover a integração de máquinas e dispositivos físicos com o espaço digital (Tang; Kuo; Lee, 2023). Ao desenvolvê-lo, permite-se a orquestração da experimentação física e computacional no ciberespaço, facilitando a automatização da descoberta química (Tao; Qi, 2019). Com isso, diminui o intervalo de espera

desde a produção de um novo produto químico no ambiente de investigação até à entrega da sua produção em massa aos utilizadores finais (Bai *et al.*, 2022).

Um exemplo muito importante de Automação de Laboratório já implementado é o Robot Scientist, que se trata um sistema robótico fisicamente implementado que faz uso de técnicas de IA para executar ciclos de experimentação científica automatizada (Whelan; King, 2004; Williams *et al.*, 2015). O Robot Scientist pode executar automaticamente ciclos de: formação de hipóteses, seleção de experimentos eficientes para discriminar hipóteses, execução de experimentos utilizando equipamentos de automação laboratorial e análise de resultados (King, 2015). A automação de laboratórios oferece uma solução para partes do fluxo de trabalho do laboratório que são rotineiras, repetitivas e de volume excessivo. Isso resulta em um trabalho de menor volume e alta qualificação para a equipe de cientistas, um subconjunto dos quais é encarregado de tratar de todo o sistema de automação (Ford; McElvania, 2020).

3.2.7 Obstáculos e oportunidades no uso das IAs

Um dos principais obstáculos nas aplicações de ML à engenharia enzimática decorre da quantidade de métodos de diferentes áreas sob uma mesma abordagem (Singh *et al.*, 2021). Bioquímicos, biólogos moleculares, matemáticos e cientistas da computação precisam chegar a um consenso para esclarecer objetivos, realizar avaliações e treinamento rigorosos e evitar problemas comuns, uso indevido de métodos e interpretações incorretas (Mazurenko; Prokop; Damborsky, 2020). Pacotes de software prontos para utilização certamente colaboram na padronização no treinamento de um algoritmo de ML para não especialistas, mas acumular todos os dados disponíveis e executar uma variedade de algoritmos de ML para selecionar o melhor preditor pode não ser a estratégia ideal (Ching *et al.*, 2018; Mazurenko; Prokop; Damborsky, 2020).

O ML promete resolver o dilema relacionada a eficiência e custo (Freeze; Kelly; Batista, 2019). É considerada uma nova revolução de paradigma para triagem virtual de alto

rendimento, previsão de estrutura e análise de desempenho de catalisadores (Zhang *et al.*, 2022). Na engenharia enzimática, o novo nível de complexidade algorítmica relacionada ao grande aumento de dados disponíveis para uso, poupará ainda mais tempo e recursos desperdiçados na validação de previsões errôneas, mas também necessitará de uma estrutura tecnológica mais sofisticada, por exemplo, um maior uso de processos de computação paralela e de treino estocástico, que já se tornaram técnicas padrão para a melhoria do treinamento de redes neurais profundas (Mazurenko; Prokop; Damborsky, 2020).

3.2.8 Aplicações na indústria farmacêutica

Diversas indústrias estão se esforçando para aprimorar seu progresso para atender às demandas e expectativas de seus clientes, utilizando diversas metodologias. A indústria farmacêutica é um campo crítico que desempenha um papel vital no salvamento de vidas (Vora *et al.*, 2023). Opera com base na inovação contínua e na adoção de novas tecnologias para enfrentar os desafios globais da saúde e responder a emergências médicas, como a recente pandemia (Vora *et al.*, 2023). A aplicação de abordagens de IA e ML é um caminho muito promissor e relativamente novo para revolucionar o processo de obtenção de medicamentos e, portanto, diminuir custos (Heifetz, 2024). A descoberta de medicamentos é o processo científico onde novas drogas e pequenas moléculas são desenvolvidos e identificados para tratar certas condições (Bakail; Ochsenbein, 2016). Ao longo da maior parcela da história da descoberta de medicamentos, foi usada uma abordagem enzimática (fechadura e chave) para identificar novas moléculas de medicamentos (Nandakumar; Dinu, 2020).

A IA revolucionou a exploração e descoberta de medicamentos de diversas formas. Algumas das principais contribuições da IA neste campo incluem o seguinte pontos: Triagem Virtual em simulações de interações químicas e prever afinidades de ligação (Bitencourt-Ferreira; Duarte da Silva; Filgueira de Azevedo, 2020); Identificação do Alvo ao analisar vários tipos de dados, como informações genéticas, proteômicos e clínicos, para identificar potenciais

alvos terapêuticos (Keum *et al.*, 2016; Wong *et al.*, 2022); Reaproveitamento de drogas ao analisar dados biomédicos em larga escala para identificar medicamentos que já existem que possam ter potencial terapêutico para diferentes doenças (Choi *et al.*, 2020; Tejera *et al.*, 2020); e Previsão de toxicidade ao prever a toxicidade dos medicamentos avaliando a estrutura química e as características dos compostos (Tabaaza *et al.*, 2023; Vora *et al.*, 2023). As abordagens baseadas em IA na pesquisa e desenvolvimento de medicamentos oferecem o potencial de otimização, agilizar a identificação e design de novos candidatos terapêuticos, levando, em última análise, a medicamentos mais eficazes e eficientes (Shah; Chavda; Soniwala, 2023).

4. TÓPICOS DE PESQUISA EMERGENTES

4.1. Análise Quantitativa das palavras-chave frequentes

Os métodos de redes de palavras-chave permitem aos pesquisadores descreverem a relação de conhecimento e elucidar a estrutura do campo (Wang, Jiamin *et al.*, 2023). Palavras-chave são os limites significativos da compreensão dos padrões e rumos dos pesquisadores que auxiliam a compreender os pontos críticos e as tendências emergentes da pesquisa ao longo do tempo (Bagdi *et al.*, 2023). A tabela 4 foi construída a partir das informações de um mapa de rede de palavras-chave e desta maneira ajuda a entender as temáticas importantes que permeiam o campo de pesquisa. Nesta tabela, 24 termos são ranqueados para entender o nível de importância e qual a sua forma de relação com outras palavras-chave demonstrado pela coluna Total Link Strength (TLS). Neste caso o TLS mostra a relação de termos, ou seja, artigos em que uma palavra-chave coexiste com outra, quanto mais artigos uma determinada palavra aparece junto de outra, mais forte a ligação entre elas e conseqüentemente compreende-se que são palavras que possuem forte relacionamento de pesquisa. Como é esperado, os termos “machine learning”(369 ocorrências), “deep learning”(88 ocorrências) e “artificial intelligence”(60 ocorrências) são os líderes desta lista, e isso se deve pelo fato de serem as abordagens de estudo, Mas um termo que pode ajudar a entender as abordagens que são

tendência neste campo é “covid-19”(47 ocorrências) na posição 4 e também na posição 8 com uma variação de nomenclatura “sars-cov-2”, isso mostra o interesse de estudo nesta área que pode ser causada principalmente pela pandemia do vírus em 2019. As ferramentas de IA podem auxiliar na triagem de pacientes e na identificação daqueles que necessitam de isolamento, a fim de evitar que espalhem o vírus para indivíduos saudáveis (Elkin; Zhu, 2022). Outro termo extremamente relevante é o que ocupa a posição 11 da tabela, “qsar” que se trata da Relação Quantitativa Estrutura-Atividade (QSAR), é uma das técnicas computacionais mais populares usadas para realizar a previsão das propriedades funcionais de compostos químicos (Samadi; Pour; Jamieson, 2021). QSAR é um modelo empírico que interpreta quantitativamente a correlação entre a estrutura química (ou propriedades relacionadas à estrutura) e a atividade biológica (ou características alvo) da substância estudada (Zhao *et al.*, 2023). Os QSAR tem por objetivo prever uma variável dependente (por exemplo, uma medida de toxicidade), fornecendo descritores químicos como variáveis independentes e, com isso, disponibilizar previsões de forma eficaz em termos de tempo e custo para produtos químicos ainda não testados (Zubrod *et al.*, 2023). E por fim o termo “molecular docking” localizado na posição 13, o docking molecular é um método de simulação teórica para analisar as interações entre moléculas (como ligantes e receptores) e prever seus modos de ligação e afinidades (Fan *et al.*, 2023). Os pesquisadores e as indústrias farmacêuticas pretendem reduzir o tempo e os custos associados ao desenvolvimento de medicamentos. O docking molecular pode ser utilizado como um filtro complexo para evidenciar apenas os candidatos a medicamentos com maior potencial (Crampon *et al.*, 2022b). Alguns termos que se encontram nesta tabela não foram analisados pois já foram citados anteriormente neste trabalho ou não possui análise pertinente para a investigação proposta.

Tabela 4

As 24 principais palavras-chave do autor na área de Inteligência Artificial em estudos enzimáticos.

Rank	Palavra-chave	Frequência	TLS	Rank	Palavra-chave	Frequência	TLS
------	---------------	------------	-----	------	---------------	------------	-----

1	machine learning	369	235	13	molecular docking	21	23
2	deep learning	88	57	14	feature selection	20	24
3	artificial intelligence	60	50	15	enzyme	19	21
4	covid-19	47	72	16	artificial neural network	18	11
5	bioinformatics	39	41	17	docking	18	26
6	support vector machine	37	36	18	molecular dynamics	17	25
7	random forest	34	41	19	biomarkers	14	6
8	sars-cov-2	33	58	20	enzymes	14	15
9	virtual screening	30	48	21	protein engineering	14	10
10	classification	24	31	22	drug discovery	13	15
11	qsar	24	33	23	synthetic biology	13	10
12	prediction	22	22	24	metabolomics	12	8

Nota: TLS: Total Link Strength

A representação ilustrada na figura 10 mostra mais incisivamente as tendências emergentes na área em estudo. É possível identificar as palavras chaves mais relevantes classificadas pelo Bibliometrix e entender qual seu crescimento durante o período de visualização. O termo com maior destaque é “prediction” acumulando aproximadamente 150 ocorrências na base de dados em análise. Conseguiu-se observar durante todo este trabalho a grande importância deste termo que envolve diversas temáticas de inteligência artificial que se relacionam ao estudo das enzimas. O gráfico demonstra que entre os anos de 2019 a 2022 foi onde ocorreu a maior evolução de ocorrências, significando que aumentou o interesse de desenvolver trabalhos referentes ao termo.

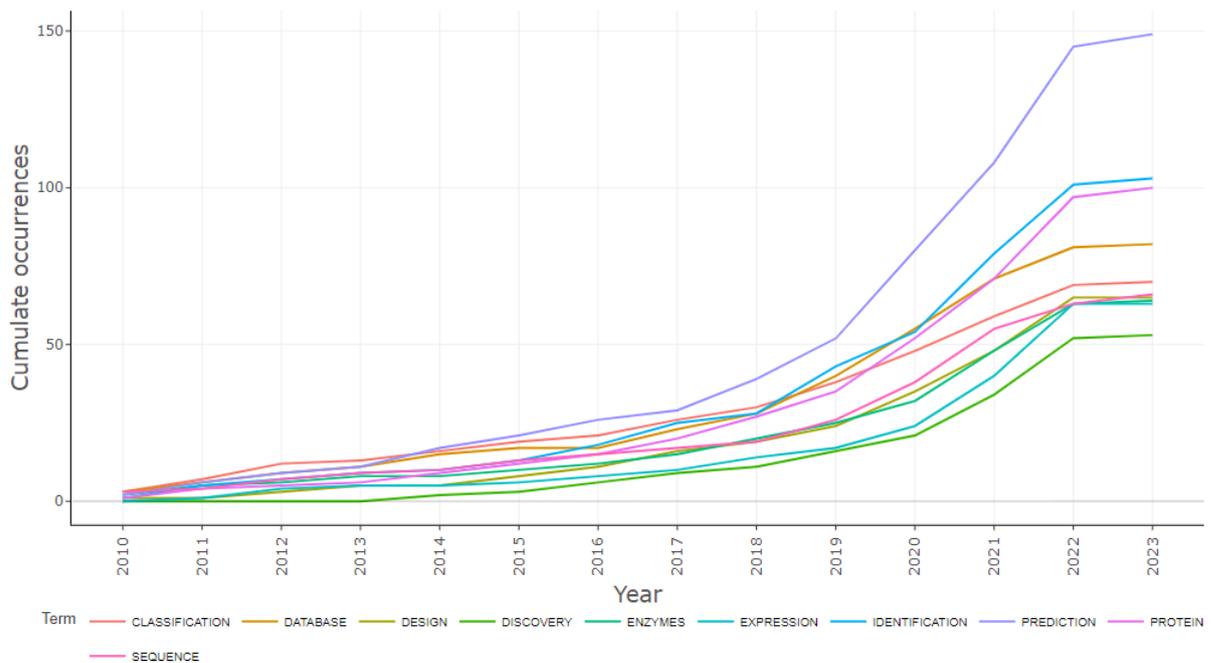


Figura 10: Tendências emergentes na área de pesquisa.

4.2. Áreas de Pesquisa

Uma forma de entender melhor como acontece a distribuição de temáticas está ilustrada na figura 11, que divide as palavras-chave de maior importância entre os 5 países mais produtivos (Estados Unidos da América, China, Índia, Alemanha e Inglaterra). Desse modo há uma compreensão de quais são as abordagens de pesquisa que mais são buscadas para investigação dentro dessas nações. Pode-se ignorar as palavras-chave derivadas de IA e observar apenas as que evidenciam os temas de estudo. Nos EUA existe dois grupos bem definidos de pesquisa que são os destacados em roxo que vão se envolver com a temática do Covid-19 (covid-19, sars-cov-2 e mutação) sendo possível observar a mesma formação na China e Índia. O outro grupo em destaque é o dos termos “substrate specificity”, “enzyme classification” e “synthetic biology” que estão todas inter-relacionadas no campo da bioquímica e biologia molecular.

A Alemanha e a Inglaterra, a partir de se cluster, é possível entender que o foco nos estudos com IA se concentram em outras áreas. Na Alemanha há uma formação de dois grupos de destaque, sendo o primeiro composto pelas palavras-chave “protein-protein interactions”,

“plants” e “phylogenetic profiling” que está ligado ao campo de utilização do perfil filogenético para prever interações proteína-proteína em plantas. Além disso, o segundo grupo com os termos “point of care”, “random forest”, “prisma” e “regression” revelam interesse de pesquisa na área médica em teste no ponto de atendimento utilizando técnicas de IA. Por fim, a Inglaterra possui somente um grande grupo formado por uma quantidade significativa de termos (“next Generation sequencing”, “analytical chemistry”, “inter-disciplinary research”, “third Generation sequencing”, “circular recycling”, “bacterial genomics” e “bacterial transcriptomics”). Estas palavras-chave fazem parte do contexto das pesquisas de degradação de plásticos e isso mostra qual eixo científico investigativo este país tem maior preocupação. A divisão geográfica visualizada pode ser explicada através de alguns fatos, principalmente a existência de interesses internos de pesquisa para uma área. A partir da figura 11 é possível ter uma visão inicial quais interesses são esses de acordo com os termos e as ligações. Outro fato é o repertório acadêmico, que em muitos casos, guia os vieses de pesquisa devido ao histórico investigativo traçado pelas nações, ou seja, uma área que já se tem costume de se trabalhar se torna algo mais interessante de investir tempo pois já se tem bases consolidadas para tal.

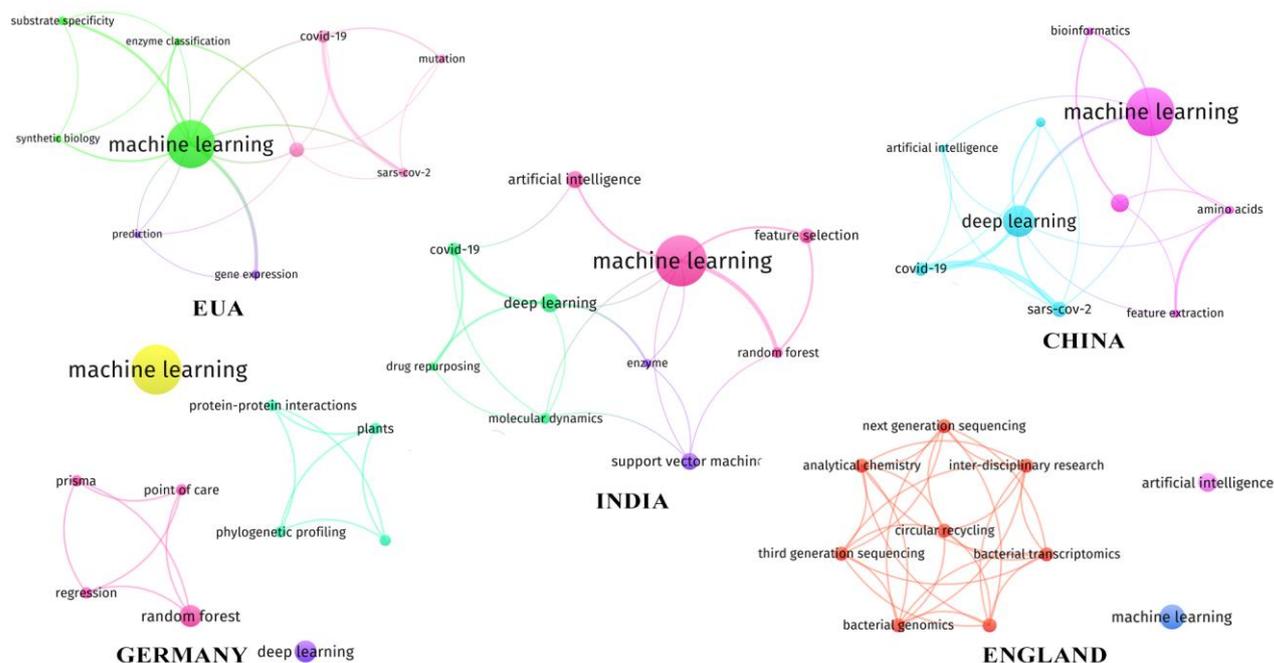
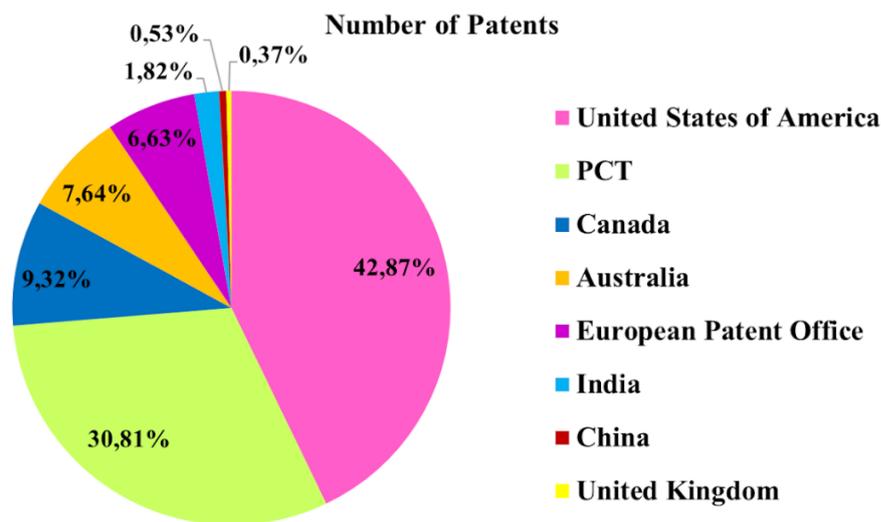


Figura 11: Mapa de visualização das palavras-chave dos 5 países com mais publicações.

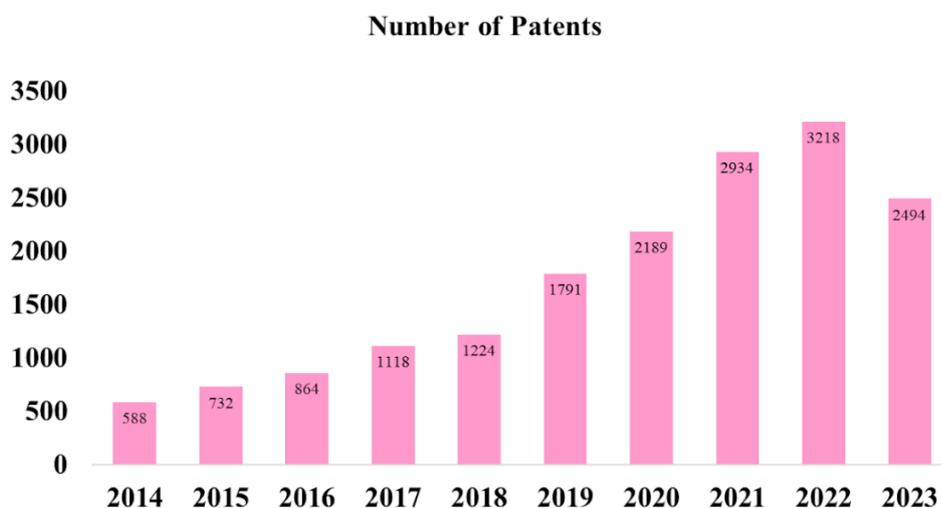
5. UMA VISÃO GERAL DAS PATENTES DE IAS NO CAMPO DA ENGENHARIA ENZIMÁTICA

A condução desta pesquisa mostrou a grande importância e crescimento alto das pesquisas no mundo relacionados a área de investigação. Ao entender o desenvolvimento científico-literário exponencial comprovado pela análise bibliométrica avançada tratada neste trabalho é possível concluir a busca ativa de conhecimento mais aprofundado. Um processo muito importante dentro das idealizações dos métodos e ferramentas computacionais de Inteligência Artificial dentro das abordagens no campo das enzimas é o patenteamento dos produtos de invenção desenvolvidos. Para verificar a quantidade de patentes acumuladas durante os anos foi utilizada a base de dados PATENTSCOPE da Organização Mundial da Propriedade Intelectual (WIPO - WORLD INTELLECTUAL PROPERTY ORGANIZATION) durante o período de janeiro de 2014 a outubro de 2023, inserindo a seguinte pesquisa ("Artificial Intelligence" OR "Machine Learning" OR "Deep Learning") AND "enzyme" no campo "todo texto" e delimitado pela língua inglesa. A pesquisa retornou 22.223 patentes registradas em todo o mundo que se relacionam com o termos utilizados sendo 6.781 do Tratado de Cooperação de Patentes (PCT - Patent Cooperation Treaty) que se trata do órgão responsável por pedidos de patente que têm validade em um grande conjunto de países.

O item (A) da figura 12 é um gráfico que mostra os países com maior número de patentes registradas nesta área. Os EUA foi o que apresentou a maior quantidade (cerca de 9.435, representando 42,87%), em segundo lugar os registros do PCT (6.781, representando 30,81%) e em terceiro lugar o Canadá (2051 patentes, representando 9,31%). A partir do item (B) da figura é possível observar que houve um crescimento grande no número de patentes registradas, onde no ano de 2014 declarou 588 patentes, número bem inferior quando comparado ao ano de 2022 que contou com a produção de 3218 patentes, número 5 vezes maior.



(A) Number of patents registered in the countries



(B) Number of patents registered between 2014 and 2023.

Figura 11: (A) Número de patentes registradas nos países com maior destaque nas pesquisas.

(B) Número de patentes registradas durante o período de 2014 a 2023.

Estes dados mostram quão promissora é esta área de estudo e o quanto de interesse os pesquisadores demonstraram em produzir e patentear ideias dentro dos estudos com enzimas. As IA fazem parte de um futuro grandioso dentro da engenharia enzimática, e os números de patentes visualizados mostram que a curva de crescimento tende a aumentar ainda mais durante os próximos anos. A partir desta análise foi possível concluir que a lista de ideias patenteáveis é bastante rica em conteúdo e quantidade e isso vai ser o pilar para a ciência não somente das enzimas, mas também para o estudo das proteínas.

6. CONCLUSÃO

Este trabalho e revisão permitiu ter uma visualização abrangente de conceitos científicos, visualizações gráficas e metodológica sobre os estudos e abordagens que envolvem as Inteligências Artificiais e todo seu contexto que deriva em Machine Learning e Deep Learning. Além disso, em destaque, realizou-se uma análise bibliométrica aprofundada que permitiu entender informações contidas no desenvolvimento literário durante o período de 2010 a 2023. Consoante à isso, várias características de relações científicas de trabalho entre grupos de autores, países, instituições e áreas de pesquisa foram identificadas e tratadas minuciosamente para estruturar e enriquecer a investigação. Com o estudo em questão, foram concluídos os seguintes pontos de discussão:

- O uso das ferramentas de Machine Learning se concentrou mais ativamente em processos de predição enzimática seja em possíveis reações, modelagem molecular, interações ou etc. A partir da quantidade de artigos analisados foi possível compreender que esta subárea foi a que demonstrou maior interesse de pesquisa.
- As maiores dificuldades de utilização das Ias são relacionadas a alta quantidade de metodologias de áreas distintas sob uma mesma abordagem onde bioquímicos, biólogos moleculares, matemáticos e cientistas da computação precisam chegar a um consenso para esclarecer objetivos, realizar avaliações e treinamento rigorosos e prevenir problemas comuns, utilização errônea de técnicas e interpretações incorretas. É imprescindível a aprendizagem de ao menos conceitos básicos de IA para o desenvolvimento harmônico, e isso é um passo inicial para resolução desses impasses citados. Ao construir uma ferramenta que envolve conceitos interdisciplinares é importante ter um elo de ligação entre conhecimentos. Por exemplo, para uma abordagem de modelagem tridimensional de uma enzima, os conceitos de estruturação esquemática, informações moleculares etc. devem ser levados em consideração como variáveis do algoritmo escolhido, por isso um conhecimento mínimo das áreas de bioquímica e biologia molecular deve ser disponibilizado a quem está desenvolvendo, e conseqüentemente conhecimentos da área de programação e IA para os especialistas em enzimas.
- O grande envolvimento dessas tecnologias com os interesses de estudo relacionado ao SARS-CoV-2 principalmente devido ao grande fato ocorrido em 2020 que culminou em uma pandemia mundial. Países como EUA e China foram os responsáveis maiores por

produzir pesquisas nesta área e maiores produtores ao todo, com 415 e 265 artigos respectivamente. Como citado, houve um grande aumento no número de dados e processamento computacional, com a pandemia, a ascensão das IAs no mundo teve um direcionamento exponencial devido a necessidade indiscutível de se obter soluções tecnológicas, inovadoras e eficazes.

- Com base nas análises do programa VOSviewer e o pacote Bibliometrix da linguagem de programação R, as áreas de pesquisas mais importantes são Bioquímica e Biologia Molecular, Química, Ciência da Computação, Biologia Computacional Matemática, Microbiologia Aplicada à Biotecnologia, Ciência Tecnológica, Farmácia e Farmacologia, Engenharia, Ciências da Vida e Biomedicina, Hereditariedade Genética, Física e Biologia Celular.
- Há uma grande quantidade de patentes criadas (22.223 registradas no PATENTSCOPE da WIPO) e isso mostra que é uma área bastante promissora sendo comprovada por este fato. Isso corrobora para um aumento dos estudos, que como visto, podem permitir uma quantidade muito alta de metodologias de pesquisa.

A exploração desta área neste trabalho proveu resultados excelentes de análise específica e ajudou a entender os rumos científicos que envolvem a temática das IAs que vão ser utilizadas direta ou indiretamente nos estudos com enzimas. Por tanto, entender mais a fundo como funcionam e quais técnicas utilizadas é altamente essencial para conseguir continuar o desenvolvimento nos próximos anos. Percebe-se grande aumento de ferramentas cada vez mais inovadoras que possibilitam a integração de conhecimentos através de várias abordagens. Isso, para o estudo com enzimas, principalmente no que tange a produção de medicamentos, é um fato extremamente atrativo. O campo da engenharia enzimática, que abrange todos os estudos aqui citados, vai ser muito beneficiado pela aplicação dessas novas tecnologias e suas futuras melhorias esperadas. A IA, principalmente ferramentas de previsão com ML, tem potencial de revolucionar este campo trazendo soluções para problemas como custos de triagem manual, período longo de análises, representação de análises complexas e dificuldade de se obter informações sobre as moléculas. O campo das ferramentas de IA ainda tem dificuldades, principalmente relacionados a quantidade de dados e o processamento sobre eles, mas é possível compreender a evolução ocorrida no número de pesquisas durante os anos, e isso mostra que a perspectiva futura é animadora. Assim, conclui-se que este trabalho representa grande passo para avaliações de interesse, riscos, oportunidades e apostas que vão estar presente mais a cada dia neste ramo da ciência.

7. REFERÊNCIAS

ABDELWAHAB, Siddig Ibrahim *et al.* Bibliometric mapping of solid lipid nanoparticles research (2012–2022) using VOSviewer. **Medicine in Novel Technology and Devices**, [s. l.], v. 17, p. 100217, 2023. Disponível em: <https://linkinghub.elsevier.com/retrieve/pii/S2590093523000127>. Acesso em: 13 de mar. de 2023.

AGWU, Okorie E. *et al.* Artificial intelligence techniques and their applications in drilling fluid engineering: A review. **Journal of Petroleum Science and Engineering**, [s. l.], v. 167, p. 300–315, 2018. Disponível em: <https://linkinghub.elsevier.com/retrieve/pii/S0920410518303255>. Acesso em: 24 de fev. de 2023.

AHMAD, Faizan. Protein stability [determination] problems. **Frontiers in Molecular Biosciences**, [s. l.], v. 9, 2022. Disponível em: <https://www.frontiersin.org/articles/10.3389/fmolb.2022.880358/full>. Acesso em: 02 de nov. de 2023.

AHUJA, Tejinder S. *et al.* Predictors of the Development of Hyperkalemia in Patients Using Angiotensin-Converting Enzyme Inhibitors. **American Journal of Nephrology**, [s. l.], v. 20, n. 4, p. 268–272, 2000. Disponível em: <https://www.karger.com/Article/FullText/13599>. Acesso em: 01 de out. de 2023.

AJJOLLI NAGARAJA, Anamya *et al.* A Machine Learning Approach for Efficient Selection of Enzyme Concentrations and Its Application for Flux Optimization. **Catalysts**, [s. l.], v. 10, n. 3, p. 291, 2020. Disponível em: <https://www.mdpi.com/2073-4344/10/3/291>. Acesso em: 01 de out. de 2023.

ALBARADEI, Somayah *et al.* Machine learning and deep learning methods that use omics data for metastasis prediction. **Computational and Structural Biotechnology Journal**, [s. l.], v. 19, p. 5008–5018, 2021. Disponível em: <https://linkinghub.elsevier.com/retrieve/pii/S200103702100386X>. Acesso em: 23 de mar. de 2023.

ALEXANDRE, Jeferson Yves Nunes Holanda *et al.* A Theoretical and Experimental Study for Enzymatic Biodiesel Production from Babassu Oil (*Orbignya sp.*) Using Eversa Lipase. **Catalysts**, [s. l.], v. 12, n. 11, p. 1322, 2022. Disponível em: <https://www.mdpi.com/2073-4344/12/11/1322>. Acesso em: 25 de fev. de 2023.

ALFORD, Rebecca F. *et al.* The Rosetta All-Atom Energy Function for Macromolecular Modeling and Design. **Journal of Chemical Theory and Computation**, [s. l.], v. 13, n. 6, p. 3031–3048, 2017. Disponível em: <https://pubs.acs.org/doi/10.1021/acs.jctc.7b00125>. Acesso em: 30 de out. de 2023.

ALJABRI, Malak *et al.* Canine impaction classification from panoramic dental radiographic images using deep learning models. **Informatics in Medicine Unlocked**, [s. l.], v. 30, p. 100918, 2022. Disponível em: <https://linkinghub.elsevier.com/retrieve/pii/S2352914822000673>. Acesso em: 13 de mar. de 2023.

ANITESCU, Cosmin *et al.* Artificial Neural Network Methods for the Solution of Second Order Boundary Value Problems. **Computers, Materials & Continua**, [s. l.], v. 59, n. 1, p. 345–359, 2019. Disponível em: <http://www.techscience.com/cmc/v59n1/27936>. Acesso em: 17 de abr. de 2023.

ARAUJO, Lilian N. *et al.* Ensemble method based on Artificial Neural Networks to estimate air pollution health risks. **Environmental Modelling & Software**, [s. l.], v. 123, p. 104567, 2020. Disponível em: <https://linkinghub.elsevier.com/retrieve/pii/S1364815219303809>. Acesso em: 17 de abr. de 2023.

ARFANUZZAMAN, Md. Harnessing artificial intelligence and big data for SDGs and prosperous urban future in South Asia. **Environmental and Sustainability Indicators**, [s. l.], v. 11, p. 100127, 2021. Disponível em: <https://linkinghub.elsevier.com/retrieve/pii/S2665972721000283>. Acesso em: 13 de mar. de 2023.

ARIA, Massimo; CUCCURULLO, Corrado. bibliometrix : An R-tool for comprehensive science mapping analysis. **Journal of Informetrics**, [s. l.], v. 11, n. 4, p. 959–975, 2017. Disponível em: <https://linkinghub.elsevier.com/retrieve/pii/S1751157717300500>. Acesso em: 16 de abr. de 2023.

AYDIN TEMEL, Fulya; CAGCAG YOLCU, Ozge; TURAN, Nurdan Gamze. Artificial intelligence and machine learning approaches in composting process: A review. **Bioresource Technology**, [s. l.], v. 370, p. 128539, 2023. Disponível em: <https://linkinghub.elsevier.com/retrieve/pii/S0960852422018727>. Acesso em: 13 de out. de 2023.

BAGDI, Tapas *et al.* Evaluation of research progress and trends on gender and renewable energy: A bibliometric analysis. **Journal of Cleaner Production**, [s. l.], v. 423, p. 138654, 2023. Disponível em: <https://linkinghub.elsevier.com/retrieve/pii/S0959652623028123>. Acesso em: 17 de nov. de 2023.

BAI, Jiaru *et al.* From Platform to Knowledge Graph: Evolution of Laboratory Automation. **JACS Au**, [s. l.], v. 2, n. 2, p. 292–309, 2022. Disponível em: <https://pubs.acs.org/doi/10.1021/jacsau.1c00438>. Acesso em: 02 de nov. de 2023.

BAIROCH, A. The SWISS-PROT protein sequence database and its supplement TrEMBL in 2000. **Nucleic Acids Research**, [s. l.], v. 28, n. 1, p. 45–48, 2000. Disponível em: <https://academic.oup.com/nar/article-lookup/doi/10.1093/nar/28.1.45>. Acesso em: 01 de nov. de 2023.

BAIROCH, A; BOECKMANN, B. The SWISS-PROT protein sequence data bank: current status. **Nucleic acids research**, [s. l.], v. 22, n. 17, p. 3578–80, 1994. Disponível em: <http://www.ncbi.nlm.nih.gov/pubmed/7937062>. Acesso em: 01 de nov. de 2023.

BAKAIL, May; OCHSENBEIN, Françoise. Targeting protein–protein interactions, a wide open field for drug design. **Comptes Rendus Chimie**, [s. l.], v. 19, n. 1–2, p. 19–27, 2016. Disponível em: <https://linkinghub.elsevier.com/retrieve/pii/S1631074816000096>. Acesso em: 02 de nov. de 2023.

BALAKRISHNAN, D.; HANEY, A. B.; MEUER, J. What a MES(s)! A bibliometric analysis of the evolution of research on multi-energy systems. **Electrical Engineering**, [s. l.], v. 98, n. 4, p. 369–374, 2016. Disponível em: <http://link.springer.com/10.1007/s00202-016-0427-9>. Acesso em: 11 de nov. de 2023.

BALOGLU, Ulas Baran *et al.* Classification of myocardial infarction with multi-lead ECG signals and deep CNN. **Pattern Recognition Letters**, [s. l.], v. 122, p. 23–30, 2019. Disponível em: <https://linkinghub.elsevier.com/retrieve/pii/S016786551930056X>. Acesso em: 27 de mai. de 2023.

BALSANO, Clara *et al.* The application of artificial intelligence in hepatology: A systematic review. **Digestive and Liver Disease**, [s. l.], v. 54, n. 3, p. 299–308, 2022. Disponível em: <https://linkinghub.elsevier.com/retrieve/pii/S1590865821003170>. Acesso em: 24 de fev. de 2023.

BARBER, Emma L. *et al.* Natural language processing with machine learning to predict outcomes after ovarian cancer surgery. **Gynecologic Oncology**, [s. l.], v. 160, n. 1, p. 182–186, 2021. Disponível em: <https://linkinghub.elsevier.com/retrieve/pii/S0090825820340129>. Acesso em: 13 de mar. de 2023.

BARREDO ARRIETA, Alejandro *et al.* Explainable Artificial Intelligence (XAI): Concepts, taxonomies, opportunities and challenges toward responsible AI. **Information Fusion**, [s. l.], v. 58, p. 82–115, 2020. Disponível em: <https://linkinghub.elsevier.com/retrieve/pii/S1566253519308103>. Acesso em: 24 de fev. de 2023.

BASHEER, Soorej M.; CHELLAPPAN, Sreeja. Enzyme Engineering. *In: BIORESOURCES AND BIOPROCESS IN BIOTECHNOLOGY*. Singapore: Springer Singapore, 2017. p. 151–168. Disponível em: http://link.springer.com/10.1007/978-981-10-4284-3_6. Acesso em: 07 de ago. de 2023.

BEDNAR, David *et al.* FireProt: Energy- and Evolution-Based Computational Design of Thermostable Multiple-Point Mutants. **PLOS Computational Biology**, [s. l.], v. 11, n. 11, p. e1004556, 2015. Disponível em: <https://dx.plos.org/10.1371/journal.pcbi.1004556>. Acesso em: 30 de out. de 2023.

BEHERA, Rajat Kumar; BALA, Pradip Kumar; RANA, Nripendra P. Creation of sustainable growth with explainable artificial intelligence: An empirical insight from consumer packaged goods retailers. **Journal of Cleaner Production**, [s. l.], v. 399, p. 136605, 2023. Disponível em: <https://linkinghub.elsevier.com/retrieve/pii/S0959652623007631>. Acesso em: 13 de Mar. de 2023.

BERG, Brandon *et al.* Cellphone-Based Hand-Held Microplate Reader for Point-of-Care Testing of Enzyme-Linked Immunosorbent Assays. **ACS Nano**, [s. l.], v. 9, n. 8, p. 7857–7866, 2015. Disponível em: <https://pubs.acs.org/doi/10.1021/acsnano.5b03203>. Acesso em: 27 de mai. de 2023.

BERG, Jens; NYSTRÖM, Kaj. A unified deep artificial neural network approach to partial differential equations in complex geometries. **Neurocomputing**, [s. l.], v. 317, p. 28–41, 2018. Disponível em: <https://linkinghub.elsevier.com/retrieve/pii/S092523121830794X>. Acesso em: 17 de abr. de 2023.

BERGMAN, A. Stevie *et al.* Representation in AI Evaluations. *In:* , 2023, New York, NY, USA. **2023 ACM Conference on Fairness, Accountability, and Transparency**. New York, NY, USA: ACM, 2023. p. 519–533. Disponível em: <https://dl.acm.org/doi/10.1145/3593013.3594019>. Acesso em: 01 de nov. de 2023.

BHATT, Harsh *et al.* State-of-the-art machine learning techniques for melanoma skin cancer detection and classification: a comprehensive review. **Intelligent Medicine**, [s. l.], 2022. Disponível em: <https://linkinghub.elsevier.com/retrieve/pii/S2667102622000754>. Acesso em: 13 de mar. de 2023.

BITAM, Said; HAMADACHE, Mabrouk; HANINI, Salah. 2D-QSAR, docking, molecular dynamics, studies of PF-07321332 analogues to identify alternative inhibitors against 3CL pro enzyme in SARS-CoV disease. **Journal of Biomolecular Structure and Dynamics**, [s. l.], v. 41, n. 14, p. 6991–7000, 2023. Disponível em:

<https://www.tandfonline.com/doi/full/10.1080/07391102.2022.2113822>. Acesso em: 20 de ago. de 2023.

BITENCOURT-FERREIRA, Gabriela; DUARTE DA SILVA, Amauri; FILGUEIRA DE AZEVEDO, Walter. Application of Machine Learning Techniques to Predict Binding Affinity for Drug Targets: A Study of Cyclin-Dependent Kinase 2. **Current Medicinal Chemistry**, [s. l.], v. 28, n. 2, p. 253–265, 2020. Disponível em: <https://www.eurekaselect.com/176353/article>. Acesso em: 02 de nov. de 2023.

BRENDEL, Matthew *et al.* Application of Deep Learning on Single-cell RNA Sequencing Data Analysis: A Review. **Genomics, Proteomics & Bioinformatics**, [s. l.], v. 20, n. 5, p. 814–835, 2022. Disponível em: <https://linkinghub.elsevier.com/retrieve/pii/S1672022922001668>. Acesso em: 13 de mar. de 2023.

BRUCKERT, Sebastian; FINZEL, Bettina; SCHMID, Ute. The Next Generation of Medical Decision Support: A Roadmap Toward Transparent Expert Companions. **Frontiers in Artificial Intelligence**, [s. l.], v. 3, 2020. Disponível em: <https://www.frontiersin.org/article/10.3389/frai.2020.507973/full>. Acesso em: 13 de nov. de 2023.

CAÑIZARES-CARMENATE, Yudith *et al.* Machine learning approach to discovery of small molecules with potential inhibitory action against vasoactive metalloproteases. **Molecular Diversity**, [s. l.], v. 26, n. 3, p. 1383–1397, 2022. Disponível em: <https://link.springer.com/10.1007/s11030-021-10260-0>. Acesso em: 05 de out. de 2023.

CAO, Chensi *et al.* Deep Learning and Its Applications in Biomedicine. **Genomics, Proteomics & Bioinformatics**, [s. l.], v. 16, n. 1, p. 17–32, 2018. Disponível em: <https://linkinghub.elsevier.com/retrieve/pii/S1672022918300020>. Acesso em: 23 de mar. de 2023.

CASADEVALL, Guillem *et al.* Estimating conformational heterogeneity of tryptophan synthase with a template-based <sc>AlphaFold2</sc> approach. **Protein Science**, [s. l.], v. 31, n. 10, 2022. Disponível em: <https://onlinelibrary.wiley.com/doi/10.1002/pro.4426>. Acesso em: 30 de out. de 2023.

CASADEVALL, Guillem; DURAN, Cristina; OSUNA, Sílvia. AlphaFold2 and Deep Learning for Elucidating Enzyme Conformational Flexibility and Its Application for Design. **JACS Au**, [s. l.], v. 3, n. 6, p. 1554–1562, 2023. Disponível em: <https://pubs.acs.org/doi/10.1021/jacsau.3c00188>. Acesso em: 30 de out. de 2023.

CATANIA, Louis J.; NICOLITZ, Ernst. Artificial Intelligence and Its Applications in Vision and Eye Care. **Advances in Ophthalmology and Optometry**, [s. l.], v. 3, n. 1, p. 21–38, 2018. Disponível em: <https://linkinghub.elsevier.com/retrieve/pii/S2452176018300015>. Acesso em: 24 de fev. de 2023.

CATUMBA, Batista Dala *et al.* Sustainability and challenges in hydrogen production: An advanced bibliometric analysis. **International Journal of Hydrogen Energy**, [s. l.], v. 48, n. 22, p. 7975–7992, 2023. Disponível em: <https://linkinghub.elsevier.com/retrieve/pii/S0360319922055343>. Acesso em: 12 de mar. De 2023.

CAVALCANTE, Francisco Thálysson Tavares *et al.* A stepwise docking and molecular dynamics approach for enzymatic biolubricant production using Lipase Eversa® Transform as a biocatalyst. **Industrial Crops and Products**, [s. l.], v. 187, p. 115450, 2022. Disponível em: <https://linkinghub.elsevier.com/retrieve/pii/S0926669022009335>. Acesso em: 25 de fev. de 2023.

CAVALCANTE, Francisco T. T. *et al.* Designing of Nanomaterials-Based Enzymatic Biosensors: Synthesis, Properties, and Applications. **Electrochem**, [s. l.], v. 2, n. 1, p. 149–184, 2021. Disponível em: <https://www.mdpi.com/2673-3293/2/1/12>. Acesso em: 25 de fev. de 2023.

CHAKRABORTY, Nilanjan *et al.* Artificial Intelligence: The road ahead for the accessibility of persons with Disability. **Materials Today: Proceedings**, [s. l.], 2021. Disponível em: <https://linkinghub.elsevier.com/retrieve/pii/S2214785321052330>. Acesso em: 13 de mar. de 2023.

CHEHILI, Hamza *et al.* Deepenz: prediction of enzyme classification by deep learning. **Indonesian Journal of Electrical Engineering and Computer Science**, [s. l.], v. 22, n. 2, p. 1108, 2021. Disponível em: <http://ijeecs.iaescore.com/index.php/IJECS/article/view/24886>. Acesso em: 01 de out. de 2023.

CHEN, Jiahui *et al.* Mutations Strengthened SARS-CoV-2 Infectivity. **Journal of Molecular Biology**, [s. l.], v. 432, n. 19, p. 5212–5226, 2020. Disponível em: <https://linkinghub.elsevier.com/retrieve/pii/S0022283620304563>. Acesso em: 14 de mai. de 2023.

CHEN, Chuangbin *et al.* Sustainability and challenges in biodiesel production from waste cooking oil: An advanced bibliometric analysis. **Energy Reports**, [s. l.], v. 7, p. 4022–4034, 2021. Disponível em: <https://linkinghub.elsevier.com/retrieve/pii/S2352484721004510>. Acesso em: 27 de mai. de 2023.

CHING, Travers *et al.* Opportunities and obstacles for deep learning in biology and medicine. **Journal of The Royal Society Interface**, [s. l.], v. 15, n. 141, p. 20170387, 2018. Disponível em: <https://royalsocietypublishing.org/doi/10.1098/rsif.2017.0387>. Acesso em: 14 de out. de 2023.

CHIU, Thomas K.F. *et al.* Systematic literature review on opportunities, challenges, and future research recommendations of artificial intelligence in education. **Computers and Education: Artificial Intelligence**, [s. l.], v. 4, p. 100118, 2023. Disponível em: <https://linkinghub.elsevier.com/retrieve/pii/S2666920X2200073X>. Acesso em: 13 de mar. de 2023.

CHOI, Yoonjung *et al.* Target-Centered Drug Repurposing Predictions of Human Angiotensin-Converting Enzyme 2 (ACE2) and Transmembrane Protease Serine Subtype 2 (TMPRSS2) Interacting Approved Drugs for Coronavirus Disease 2019 (COVID-19) Treatment through a Drug-Target Interacti. **Viruses**, [s. l.], v. 12, n. 11, p. 1325, 2020. Disponível em: <https://www.mdpi.com/1999-4915/12/11/1325>. Acesso em: 02 de nov. de 2023.

CHOW, Daniel S. *et al.* Updates on Deep Learning and Glioma. **Neuroimaging Clinics of North America**, [s. l.], v. 30, n. 4, p. 493–503, 2020. Disponível em: <https://linkinghub.elsevier.com/retrieve/pii/S1052514920300526>. Acesso em: 23 de mar. de 2023.

CHUNGYOUN, Michael F.; GRAY, Jeffrey J. AI models for protein design are driving antibody engineering. **Current Opinion in Biomedical Engineering**, [s. l.], v. 28, p. 100473, 2023. Disponível em: <https://linkinghub.elsevier.com/retrieve/pii/S2468451123000296>. Acesso em: 05 de out. de 2023.

CONCU, Riccardo *et al.* Mozart, a Qsar Multi-Target Web Based Tool to Predict Multiple Drug-Enzyme Interactions. **SSRN Electronic Journal**, [s. l.], 2022. Disponível em: <https://www.ssrn.com/abstract=4057339>. Acesso em: 20 de ago. de 2023.

CONG, Yong *et al.* Prediction of novel and selective TNF-alpha converting enzyme (TACE) inhibitors and characterization of correlative molecular descriptors by machine learning approaches.

Journal of Molecular Graphics and Modelling, [s. l.], v. 28, n. 3, p. 236–244, 2009. Disponível em: <https://linkinghub.elsevier.com/retrieve/pii/S1093326309000874>. Acesso em: 25 de fev. de 2023.

CRAMPON, Kevin *et al.* Machine-learning methods for ligand–protein molecular docking. **Drug Discovery Today**, [s. l.], v. 27, n. 1, p. 151–164, 2022a. Disponível em: <https://linkinghub.elsevier.com/retrieve/pii/S1359644621003974>. Acesso em: 17 de nov. de 2023.

CSOSZ, Eva *et al.* Substrate Preference of Transglutaminase 2 Revealed by Logistic Regression Analysis and Intrinsic Disorder Examination. **Journal of Molecular Biology**, [s. l.], v. 383, n. 2, p. 390–402, 2008. Disponível em: <https://linkinghub.elsevier.com/retrieve/pii/S002228360801019X>. Acesso em: 15 de nov. de 2023.

DAMBORSKY, Jiri; BREZOVSKY, Jan. Computational tools for designing and engineering enzymes. **Current Opinion in Chemical Biology**, [s. l.], v. 19, p. 8–16, 2014. Disponível em: <https://linkinghub.elsevier.com/retrieve/pii/S1367593113002354>. Acesso em: 30 de out. de 2023.

DAMULIENĖ, Vaida *et al.* The Influence of Enzymatic Hydrolysis on Bee Pollen Antioxidant and Antibacterial Activities. **Foods**, [s. l.], v. 12, n. 19, p. 3582, 2023. Disponível em: <https://www.mdpi.com/2304-8158/12/19/3582>. Acesso em: 15 de nov. de 2023.

DAUPARAS, J. *et al.* Robust deep learning–based protein sequence design using ProteinMPNN. **Science**, [s. l.], v. 378, n. 6615, p. 49–56, 2022. Disponível em: <https://www.science.org/doi/10.1126/science.add2187>. Acesso em: 30 de out. de 2023.

DE MENEZES, Luiz Henrique Sales *et al.* The Application of Chemometric Methods in the Production of Enzymes Through Solid State Fermentation Uses the Artificial Neural Network—a Review. **BioEnergy Research**, [s. l.], v. 16, n. 1, p. 279–288, 2023. Disponível em: <https://link.springer.com/10.1007/s12155-022-10462-w>. Acesso em: 20 de ago. de 2023.

DE SOUSA, Isamayra Germano *et al.* Renewable processes of synthesis of biolubricants catalyzed by lipases. **Journal of Environmental Chemical Engineering**, [s. l.], v. 11, n. 1, p. 109006, 2023. Disponível em: <https://linkinghub.elsevier.com/retrieve/pii/S2213343722018796>. Acesso em: de 2023. Acesso em: 25 de fev. de 2023.

DEHOUCK, Yves *et al.* Fast and accurate predictions of protein stability changes upon mutations using statistical potentials and neural networks: PoPMuSiC-2.0. **Bioinformatics**, [s. l.], v. 25, n. 19, p. 2537–2543, 2009. Disponível em: <https://academic.oup.com/bioinformatics/article/25/19/2537/180561>. Acesso em: 30 de out. de 2023.

DEIG, Christopher R.; KANWAR, Aasheesh; THOMPSON, Reid F. Artificial Intelligence in Radiation Oncology. **Hematology/Oncology Clinics of North America**, [s. l.], v. 33, n. 6, p. 1095–1104, 2019. Disponível em: <https://linkinghub.elsevier.com/retrieve/pii/S0889858819300966>. Acesso em: 13 de mar. de 2023.

DERAT, Etienne; KAMERLIN, Shina Caroline Lynn. Computational Advances in Protein Engineering and Enzyme Design. **The Journal of Physical Chemistry B**, [s. l.], v. 126, n. 13, p. 2449–2451, 2022. Disponível em: <https://pubs.acs.org/doi/10.1021/acs.jpccb.2c01198>. Acesso em: 07 de ago. de 2023.

DERVIŞ, Hamid. Bibliometric Analysis using Bibliometrix an R Package. **Journal of Scientometric Research**, [s. l.], v. 8, n. 3, p. 156–160, 2020. Disponível em: <http://www.jscires.org/article/326>. Acesso em: 14 de mai. de 2023.

DHARMIK, Rajesh C.; BAWANKAR, Bhushan U. Design Challenges for Machine/Deep Learning Algorithms. *In: MACHINE LEARNING FOR VLSI CHIP DESIGN*. [S. l.]: Wiley, 2023. p. 195–209. Disponível em: <https://onlinelibrary.wiley.com/doi/10.1002/9781119910497.ch13>. Acesso em: 01 de nov. de 2023.

DIXIT, Mandeep; CHHABRA, Deepak; SHUKLA, Pratyosh. Optimization of endoglucanase-lipase-amylase enzyme consortium from *Thermomyces lanuginosus* VAPS25 using Multi-Objective genetic algorithm and their bio-deinking applications. **Bioresource Technology**, [s. l.], v. 370, p. 128467, 2023. Disponível em: <https://linkinghub.elsevier.com/retrieve/pii/S0960852422018004>. Acesso em: 20 de ago. de 2023.

DOBSON, Paul D.; DOIG, Andrew J. Predicting Enzyme Class From Protein Structure Without Alignments. **Journal of Molecular Biology**, [s. l.], v. 345, n. 1, p. 187–199, 2005. Disponível em: <https://linkinghub.elsevier.com/retrieve/pii/S0022283604013166>. Acesso em: 14 de ago. de 2023.

DONATI, G. Increased prevalence of fatty liver in arterial hypertensive patients with normal liver enzymes: role of insulin resistance. **Gut**, [s. l.], v. 53, n. 7, p. 1020–1023, 2004. Disponível em: <https://gut.bmj.com/lookup/doi/10.1136/gut.2003.027086>. Acesso em: 01 de ago. de 2023.

DONTHU, Naveen *et al.* How to conduct a bibliometric analysis: An overview and guidelines. **Journal of Business Research**, [s. l.], v. 133, p. 285–296, 2021. Disponível em: <https://linkinghub.elsevier.com/retrieve/pii/S0148296321003155>. Acesso em: 17 de abr. de 2023.

DOW, Eliot R. *et al.* From Data to Deployment. **Ophthalmology**, [s. l.], v. 129, n. 5, p. e43–e59, 2022. Disponível em: <https://linkinghub.elsevier.com/retrieve/pii/S0161642022000021>. Acesso em: 24 de fev. de 2023.

DRISCOLL, Lewis; DE LA TORRE, Sebastián; GOMEZ-RUIZ, Jose Antonio. Feature-based lithium-ion battery state of health estimation with artificial neural networks. **Journal of Energy Storage**, [s. l.], v. 50, p. 104584, 2022. Disponível em: <https://linkinghub.elsevier.com/retrieve/pii/S2352152X22006004>. Acesso em: 17 de abr. de 2023.

DUAN, Yanqing; EDWARDS, John S.; DWIVEDI, Yogesh K. Artificial intelligence for decision making in the era of Big Data – evolution, challenges and research agenda. **International Journal of Information Management**, [s. l.], v. 48, p. 63–71, 2019. Disponível em: <https://linkinghub.elsevier.com/retrieve/pii/S0268401219300581>. Acesso em: 13 de mar. de 2023.

DUHAN, Naveen; NORTON, Jeanette M; KAUNDAL, Rakesh. deepNEC: a novel alignment-free tool for the identification and classification of nitrogen biochemical network-related enzymes using deep learning. **Briefings in Bioinformatics**, [s. l.], v. 23, n. 3, 2022. Disponível em: <https://academic.oup.com/bib/article/doi/10.1093/bib/bbac071/6553605>. Acesso em: 01 de out. de 2023.

EBRAHIMI, Mansour *et al.* Prediction of Thermostability from Amino Acid Attributes by Combination of Clustering with Attribute Weighting: A New Vista in Engineering Enzymes. **PLoS ONE**, [s. l.], v. 6, n. 8, p. e23146, 2011. Disponível em: <https://dx.plos.org/10.1371/journal.pone.0023146>. Acesso em: 01 de out. de 2023.

EBRAHIMI, Mansour; EBRAHIMIE, Esmail. Sequence-Based Prediction of Enzyme Thermostability Through Bioinformatics Algorithms. **Current Bioinformatics**, [s. l.], v. 5, n. 3, p. 195–203, 2010. Disponível em: <http://www.eurekaselect.com/openurl/content.php?genre=article&iissn=1574-8936&volume=5&issue=3&spage=195>. Acesso em: 01 de out. de 2023.

ELKIN, Magdalyn E.; ZHU, Xingquan. A machine learning study of COVID-19 serology and molecular tests and predictions. **Smart Health**, [s. l.], v. 26, p. 100331, 2022. Disponível em: <https://linkinghub.elsevier.com/retrieve/pii/S2352648322000654>. Acesso em: 18 de out. de 2023.

ELLEGAARD, Ole; WALLIN, Johan A. The bibliometric analysis of scholarly production: How great is the impact?. **Scientometrics**, [s. l.], v. 105, n. 3, p. 1809–1831, 2015. Disponível em: <http://link.springer.com/10.1007/s11192-015-1645-z>. Acesso em: 16 de abr. de 2023.

FÁGÁIN, Ciarán Ó. Understanding and increasing protein stability. **Biochimica et Biophysica Acta (BBA) - Protein Structure and Molecular Enzymology**, [s. l.], v. 1252, n. 1, p. 1–14, 1995. Disponível em: <https://linkinghub.elsevier.com/retrieve/pii/016748389500133F>. Acesso em: 30 de out. de 2023.

FAHIMNIA, Behnam; SARKIS, Joseph; DAVARZANI, Hoda. Green supply chain management: A review and bibliometric analysis. **International Journal of Production Economics**, [s. l.], v. 162, p. 101–114, 2015. Disponível em: <https://linkinghub.elsevier.com/retrieve/pii/S0925527315000067>. Acesso em: 04 de mai. de 2023.

FAN, Xiankang *et al.* Preparation and characterization of duck liver-derived antioxidant peptides based on LC-MS/MS, molecular docking, and machine learning. **LWT**, [s. l.], v. 175, p. 114479, 2023. Disponível em: <https://linkinghub.elsevier.com/retrieve/pii/S0023643823000579>. Acesso em: 18 de out. de 2023.

FAN, Hu. Research on innovation and application of 5G using artificial intelligence-based image and speech recognition technologies. **Journal of King Saud University - Science**, [s. l.], v. 35, n. 4, p. 102626, 2023. Disponível em: <https://linkinghub.elsevier.com/retrieve/pii/S1018364723000885>. Acesso em: 13 de mar. de 2023.

FEEHAN, RYAN; MONTEZANO, DANIEL ; SLUSKY, Joanna S G. Machine learning for enzyme engineering, selection and design. **Protein Engineering, Design, and Selection**, [s. l.], v. 34, 2021. Disponível em: http://fdslive.oup.com/www.oup.com/pdf/production_in_progress.pdf. Acesso em: 14 de ago. de 2023.

FERNANDES, Clara Dourado *et al.* Fungal biosynthesis of lignin-modifying enzymes from pulp wash and *Luffa cylindrica* for azo dye RB5 biodecolorization using modeling by response surface methodology and artificial neural network. **Journal of Hazardous Materials**, [s. l.], v. 399, p. 123094, 2020. Disponível em: <https://linkinghub.elsevier.com/retrieve/pii/S0304389420310839>. Acesso em: 25 de fev. de 2023.

FERNANDEZ, Michael; CABALLERO, Julio. Genetic Algorithm Optimization of Bayesian-Regularized Artificial Neural Networks in Drug Design. *In: ARTIFICIAL NEURAL NETWORK FOR DRUG DESIGN, DELIVERY AND DISPOSITION*. [S. l.]: Elsevier, 2016. p. 83–102. Disponível em: <https://linkinghub.elsevier.com/retrieve/pii/B9780128015599000041>. Acesso em: 20 de ago. de 2023.

FERNÁNDEZ, Michael; CABALLERO, Julio. Modeling of activity of cyclic urea HIV-1 protease inhibitors using regularized-artificial neural networks. **Bioorganic & Medicinal Chemistry**, [s. l.], v. 14, n. 1, p. 280–294, 2006. Disponível em: <https://linkinghub.elsevier.com/retrieve/pii/S0968089605007765>. Acesso em: 01 de out. de 2023.

FERREIRA MOTA, Gabrielly *et al.* Biodiesel production from microalgae using lipase-based catalysts: Current challenges and prospects. **Algal Research**, [s. l.], v. 62, p. 102616, 2022. Disponível em: <https://linkinghub.elsevier.com/retrieve/pii/S2211926421004355>. Acesso em: 25 de fev. de 2023.

FOLKMAN, Lukas *et al.* EASE-MM: Sequence-Based Prediction of Mutation-Induced Stability Changes with Feature-Based Multiple Models. **Journal of Molecular Biology**, [s. l.], v. 428, n. 6, p. 1394–1405, 2016. Disponível em: <https://linkinghub.elsevier.com/retrieve/pii/S0022283616000310>. Acesso em: 30 de out. de 2023.

FORD, Bradley A.; MCELVANIA, Erin. Machine Learning Takes Laboratory Automation to the Next Level. **Journal of Clinical Microbiology**, [s. l.], v. 58, n. 4, 2020. Disponível em: <https://journals.asm.org/doi/10.1128/JCM.00012-20>. Acesso em: 02 de nov. de 2023.

FREEZE, Jessica G.; KELLY, H. Ray; BATISTA, Victor S. Search for Catalysts by Inverse Design: Artificial Intelligence, Mountain Climbers, and Alchemists. **Chemical Reviews**, [s. l.], v. 119, n. 11, p. 6595–6612, 2019. Disponível em: <https://pubs.acs.org/doi/10.1021/acs.chemrev.8b00759>. Acesso em: 14 de out. de 2023.

GAO, Yu-Fei *et al.* Prediction of active sites of enzymes by maximum relevance minimum redundancy (mRMR) feature selection. **Mol. BioSyst.**, [s. l.], v. 9, n. 1, p. 61–69, 2013. Disponível em: <http://xlink.rsc.org/?DOI=C2MB25327E>. Acesso em: 01 de out. de 2023.

GE, Feiyin *et al.* Artificial Intelligence Aided Lipase Production and Engineering for Enzymatic Performance Improvement. **Journal of Agricultural and Food Chemistry**, [s. l.], v. 71, n. 41, p. 14911–14930, 2023. Disponível em: <https://pubs.acs.org/doi/10.1021/acs.jafc.3c05029>. Acesso em: 01 de nov. de 2023.

GHOBIAN, B. *et al.* Diesel engine performance and exhaust emission analysis using waste cooking biodiesel fuel with an artificial neural network. **Renewable Energy**, [s. l.], v. 34, n. 4, p. 976–982, 2009. Disponível em: <https://linkinghub.elsevier.com/retrieve/pii/S0960148108003108>. Acesso em: 17 de abr. de 2023.

GHOSE, Anamika *et al.* Optimization of laccase enzyme extraction from spent mushroom waste of *Pleurotus florida* through ANN-PSO modeling: An ecofriendly and economical approach. **Environmental Research**, [s. l.], v. 222, p. 115345, 2023. Disponível em: <https://linkinghub.elsevier.com/retrieve/pii/S0013935123001378>. Acesso em: 20 de ago. de 2023.

GOLINELLI-PIMPANEAU, Béatrice. Prediction of the Iron–Sulfur Binding Sites in Proteins Using the Highly Accurate Three-Dimensional Models Calculated by AlphaFold and RoseTTAFold. **Inorganics**, [s. l.], v. 10, n. 1, p. 2, 2021. Disponível em: <https://www.mdpi.com/2304-6740/10/1/2>. Acesso em: 30 de out. de 2023.

GOMES, Eleni *et al.* Enzimas termoestáveis: fontes, produção e aplicação industrial. **Química Nova**, [s. l.], v. 30, n. 1, p. 136–145, 2007. Disponível em: http://www.scielo.br/scielo.php?script=sci_arttext&pid=S0100-40422007000100025&lng=pt&nrm=iso&tlng=pt. Acesso em: 02 de nov. de 2023.

GUARASCIO, Massimo; MANCO, Giuseppe; RITACCO, Ettore. Deep Learning. *In: ENCYCLOPEDIA OF BIOINFORMATICS AND COMPUTATIONAL BIOLOGY*. [s. l.]: Elsevier, 2019. p. 634–647. Disponível em: <https://linkinghub.elsevier.com/retrieve/pii/B978012809633820352X>. Acesso em: 23 de mar. de 2023.

GUIMARÃES, Cláudia Elisa César *et al.* Sustainable bioethanol production from first- and second-generation sugar-based feedstocks: Advanced bibliometric analysis. **Bioresource Technology Reports**, [s. l.], v. 23, p. 101543, 2023. Disponível em: <https://linkinghub.elsevier.com/retrieve/pii/S2589014X23002141>. Acesso em: 13 de nov. de 2023.

GÜNAY, M. Erdem *et al.* Simultaneous modeling of enzyme production and biomass growth in recombinant *Escherichia coli* using artificial neural networks. **Biochemical Engineering Journal**, [s. l.], v. 42, n. 3, p. 329–335, 2008. Disponível em: <https://linkinghub.elsevier.com/retrieve/pii/S1369703X0800257X>. Acesso em: 13 de nov. de 2023.

GUTTERIDGE, Alex; BARTLETT, Gail J; THORNTON, Janet M. Using A Neural Network and Spatial Clustering to Predict the Location of Active Sites in Enzymes. **Journal of Molecular Biology**, [s. l.], v. 330, n. 4, p. 719–734, 2003. Disponível em: <https://linkinghub.elsevier.com/retrieve/pii/S0022283603005151>. Acesso em: 14 de ago. de 2023.

HABERMAN, Cory P. *et al.* The sensitivity of repeat and near repeat analysis to geocoding algorithms. **Journal of Criminal Justice**, [s. l.], v. 73, p. 101721, 2021. Disponível em: <https://linkinghub.elsevier.com/retrieve/pii/S0047235220302154>. Acesso em: 16 de abr. de 2023.

HADDAJI, Achref; AYED, Samiha; FOURATI, Lamia Chaari. Artificial Intelligence techniques to mitigate cyber-attacks within vehicular networks: Survey. **Computers and Electrical Engineering**, [s. l.], v. 104, p. 108460, 2022. Disponível em: <https://linkinghub.elsevier.com/retrieve/pii/S0045790622006759>. Acesso em: 13 de mar. de 2023.

HALEEM, Abid; JAVAID, Mohd; KHAN, Ibrahim Haleem. Current status and applications of Artificial Intelligence (AI) in medical field: An overview. **Current Medicine Research and Practice**, [s. l.], v. 9, n. 6, p. 231–237, 2019. Disponível em: <https://linkinghub.elsevier.com/retrieve/pii/S235208171930193X>. Acesso em: 13 de mar. de 2023.

HAMMER, Alexander J. S. *et al.* Chemputation and the Standardization of Chemical Informatics. **JACS Au**, [s. l.], v. 1, n. 10, p. 1572–1587, 2021. Disponível em: <https://pubs.acs.org/doi/10.1021/jacsau.1c00303>. Acesso em: 02 de nov. de 2023.

HAQUE, Shafiul *et al.* Artificial Intelligence vs. Statistical Modeling and Optimization of Continuous Bead Milling Process for Bacterial Cell Lysis. **Frontiers in Microbiology**, [s. l.], v. 7, 2016. Disponível em: <http://journal.frontiersin.org/article/10.3389/fmicb.2016.01852/full>. Acesso em: 05 de ou. de 2023.

HEIFETZ, Alexander. Accelerating COVID-19 Drug Discovery with High-Performance Computing. *In: [S. l.: s. n.]*, 2024. p. 405–411. Disponível em: https://link.springer.com/10.1007/978-1-0716-3449-3_19. Acesso em: 02 de nov. de 2023.

HOGAN, James M.; PEUT, Timothy. Large Scale Read Classification for Next Generation Sequencing. **Procedia Computer Science**, [s. l.], v. 29, p. 2003–2012, 2014. Disponível em: <https://linkinghub.elsevier.com/retrieve/pii/S1877050914003615>. Acesso em: 20 de ago. de 2023.

HOSSEINI, Seyed Nezamedin; JAVIDANBARDAN, Amin; KHATAMI, Maryam. Accurate and cost-effective prediction of HBsAg titer in industrial scale fermentation process of recombinant *Pichia pastoris* by using neural network based soft sensor. **Biotechnology and Applied Biochemistry**, [s. l.], v. 66, n. 4, p. 681–689, 2019. Disponível em: <https://iubmb.onlinelibrary.wiley.com/doi/10.1002/bab.1785>. Acesso em: 01 de out. de 2023.

HUANG, Pengxinyue *et al.* A bibliometric analysis of global research hotspots and progress on microplastics in soil–plant systems. **Environmental Pollution**, [s. l.], p. 122890, 2023. Disponível em: <https://linkinghub.elsevier.com/retrieve/pii/S0269749123018924>. Acesso em: 01 de out. de 2023.

HUANG, Ailan; LU, Fuping; LIU, Fufeng. Discrimination of psychrophilic enzymes using machine learning algorithms with amino acid composition descriptor. **Frontiers in Microbiology**, [s. l.], v. 14, 2023. Disponível em: <https://www.frontiersin.org/articles/10.3389/fmicb.2023.1130594/full>. Acesso em: 13 de nov. de 2023.

IYER, Padma V.; ANANTHANARAYAN, Laxmi. Enzyme stability and stabilization—Aqueous and non-aqueous environment. **Process Biochemistry**, [s. l.], v. 43, n. 10, p. 1019–1032, 2008. Disponível em: <https://linkinghub.elsevier.com/retrieve/pii/S1359511308001876>. Acesso em: 30 de out. de 2023.

JANG, Woo Dae *et al.* Applications of artificial intelligence to enzyme and pathway design for metabolic engineering. **Current Opinion in Biotechnology**, [s. l.], v. 73, p. 101–107, 2022a. Disponível em: <https://linkinghub.elsevier.com/retrieve/pii/S0958166921001361>. Acesso em: 05 de nov. de 2023.

JAREMKO, Jacob L. *et al.* Canadian Association of Radiologists White Paper on Ethical and Legal Issues Related to Artificial Intelligence in Radiology. **Canadian Association of Radiologists Journal**, [s. l.], v. 70, n. 2, p. 107–118, 2019. Disponível em: <http://journals.sagepub.com/doi/10.1016/j.carj.2019.03.001>. Acesso em: 24 de fev. de 2023.

JIANG, Xiangao *et al.* Towards an Artificial Intelligence Framework for Data-Driven Prediction of Coronavirus Clinical Severity. **Computers, Materials & Continua**, [s. l.], v. 62, n. 3, p. 537–551, 2020. Disponível em: <https://www.techscience.com/cmc/v63n1/38464>. Acesso em: 27 de mai. de 2023.

JIANG, Yaoyukun; RAN, Xinchun; YANG, Zhongyue J. Data-driven enzyme engineering to identify function-enhancing enzymes. **Protein Engineering, Design and Selection**, [s. l.], v. 36, 2023. Disponível em: <https://academic.oup.com/peds/article/doi/10.1093/protein/gzac009/6754785>. Acesso em: 14 de ago. de 2023.

JIAO, Zengtao *et al.* Application of big data and artificial intelligence in epidemic surveillance and containment. **Intelligent Medicine**, [s. l.], v. 3, n. 1, p. 36–43, 2023. Disponível em: <https://linkinghub.elsevier.com/retrieve/pii/S266710262200078X>. Acesso em: 30 de out. de 2023.

JIAO, Yanqi; XING, Yichen; SUN, Yao. Impact of E484Q and L452R Mutations on Structure and Binding Behavior of SARS-CoV-2 B.1.617.1 Using Deep Learning AlphaFold2, Molecular Docking and Dynamics Simulation. **International Journal of Molecular Sciences**, [s. l.], v. 24, n. 14, p. 11564, 2023. Disponível em: <https://www.mdpi.com/1422-0067/24/14/11564>. Acesso em: 13 de mar. de 2023.

JIMMA, Bahiru Legesse. Artificial intelligence in healthcare: A bibliometric analysis. **Telematics and Informatics Reports**, [s. l.], v. 9, p. 100041, 2023. Disponível em: <https://linkinghub.elsevier.com/retrieve/pii/S2772503023000014>. Acesso em: 13 de mar. de 2023.

KACZMAREK, Iwona *et al.* A machine learning approach for integration of spatial development plans based on natural language processing. **Sustainable Cities and Society**, [s. l.], v. 76, p. 103479, 2022. Disponível em: <https://linkinghub.elsevier.com/retrieve/pii/S2210670721007460>. Acesso em: 13 de mar. de 2023.

KADHIM, Zahraa Saddi; ABDULLAH, Hasanen S.; GHATHWAN, Khalil I. Automatically Avoiding Overfitting in Deep Neural Networks by Using Hyper-Parameters Optimization Methods. **International Journal of Online and Biomedical Engineering (iJOE)**, [s. l.], v. 19, n. 05, p. 146–162, 2023. Disponível em: <https://online-journals.org/index.php/i-joe/article/view/38153>. Acesso em: 01 de nov. de 2023.

KAO, Hsi-Wen *et al.* Robust Design of Effective Allosteric Activators for Rsp5 E3 Ligase Using the Machine Learning Tool ProteinMPNN. **ACS Synthetic Biology**, [s. l.], v. 12, n. 8, p. 2310–2319, 2023. Disponível em: <https://pubs.acs.org/doi/10.1021/acssynbio.3c00042>. Acesso em: 30 de out. de 2023.

KAR, Arpan Kumar; CHOUDHARY, Shweta Kumari; SINGH, Vinay Kumar. How can artificial intelligence impact sustainability: A systematic literature review. **Journal of Cleaner Production**, [s. l.], v. 376, p. 134120, 2022. Disponível em: <https://linkinghub.elsevier.com/retrieve/pii/S0959652622036927>. Acesso em: 13 de mar. de 2023.

KARP, Peter D.; PALEY, Suzanne; ZHU, Jingchun. Database verification studies of SWISS-PROT and GenBank. **Bioinformatics**, [s. l.], v. 17, n. 6, p. 526–532, 2001. Disponível em: <https://academic.oup.com/bioinformatics/article/17/6/526/272368>. Acesso em: 01 de nov. de 2023.

KATO, Tsuyoshi; NAGANO, Nozomi. Discriminative structural approaches for enzyme active-site prediction. **BMC Bioinformatics**, [s. l.], v. 12, n. S1, p. S49, 2011. Disponível em: <https://bmcbioinformatics.biomedcentral.com/articles/10.1186/1471-2105-12-S1-S49>. Acesso em: 14 de ago. de 2023.

KAUR, Manpreet; RATTAN, Dhavleesh. A systematic literature review on the use of machine learning in code clone research. **Computer Science Review**, [s. l.], v. 47, p. 100528, 2023. Disponível em: <https://linkinghub.elsevier.com/retrieve/pii/S1574013722000624>. Acesso em: 24 de fev. de 2023.

KAUTSAR, Satria A *et al.* MIBiG 2.0: a repository for biosynthetic gene clusters of known function. **Nucleic Acids Research**, [s. l.], 2019. Disponível em: <https://academic.oup.com/nar/advance-article/doi/10.1093/nar/gkz882/5587631>. Acesso em: 01 de nov. de 2023.

KEUM, Jongsoo *et al.* Prediction of compound-target interactions of natural products using large-scale drug and protein information. **BMC Bioinformatics**, [s. l.], v. 17, n. S6, p. 219, 2016. Disponível em: <http://bmcbioinformatics.biomedcentral.com/articles/10.1186/s12859-016-1081-y>. Acesso em: 02 de nov. de 2023.

KHAN, Zaheer Ullah; HAYAT, Maqsood; KHAN, Muazzam Ali. Discrimination of acidic and alkaline enzyme using Chou's pseudo amino acid composition in conjunction with probabilistic neural network model. **Journal of Theoretical Biology**, [s. l.], v. 365, p. 197–203, 2015. Disponível

em: <https://linkinghub.elsevier.com/retrieve/pii/S0022519314006031>. Acesso em: 01 de out. de 2023.

KHANAL, Samir Kumar; TARAFDAR, Ayon; YOU, Siming. Artificial intelligence and machine learning for smart bioprocesses. **Bioresource Technology**, [s. l.], v. 375, p. 128826, 2023. Disponível em: <https://linkinghub.elsevier.com/retrieve/pii/S0960852423002523>. Acesso em: 23 de mar. de 2023.

KING, Ross D. Automating Chemistry and Biology Using Robot Scientists. *In:* , 2015. (S. Holldobler, S; Krotzsch, M; Penaloza, R; Rudolph, Org.) **Lecture Notes in Artificial Intelligence**. [S. l.]: ADVANCES IN INTELLIGENCE ARTIFICIAL, 2015. p. XIV–XV. Acesso em: 02 de nov. de 2023.

KIRCHMAIR, Johannes *et al.* Computational Prediction of Metabolism: Sites, Products, SAR, P450 Enzyme Dynamics, and Mechanisms. **Journal of Chemical Information and Modeling**, [s. l.], v. 52, n. 3, p. 617–648, 2012. Disponível em: <https://pubs.acs.org/doi/10.1021/ci200542m>. Acesso em: 01 de out. de 2023.

KIRCHMAIR, Johannes *et al.* FAst MEtabolizer (FAME): A Rapid and Accurate Predictor of Sites of Metabolism in Multiple Species by Endogenous Enzymes. **Journal of Chemical Information and Modeling**, [s. l.], v. 53, n. 11, p. 2896–2907, 2013. Disponível em: <https://pubs.acs.org/doi/10.1021/ci400503s>. Acesso em: 27 de mai. de 2023.

KOMEILIBIRJANDI, Ali *et al.* Thermal conductivity prediction of nanofluids containing CuO nanoparticles by using correlation and artificial neural network. **Journal of Thermal Analysis and Calorimetry**, [s. l.], v. 139, n. 4, p. 2679–2689, 2020. Disponível em: <http://link.springer.com/10.1007/s10973-019-08838-w>. Acesso em: 17 de abr. de 2023.

KORKEGIAN, Aaron *et al.* Computational Thermostabilization of an Enzyme. **Science**, [s. l.], v. 308, n. 5723, p. 857–860, 2005. Disponível em: <https://www.science.org/doi/10.1126/science.1107387>. Acesso em: 02 de nov. de 2023.

KOTSIANTIS, S. B.; ZAHARAKIS, I. D.; PINTELAS, P. E. Machine learning: a review of classification and combining techniques. **Artificial Intelligence Review**, [s. l.], v. 26, n. 3, p. 159–190, 2006. Disponível em: <http://link.springer.com/10.1007/s10462-007-9052-3>. Acesso em: 17 de abr. de 2023.

KROLL, Alexander *et al.* A general model to predict small molecule substrates of enzymes based on machine and deep learning. **Nature Communications**, [s. l.], v. 14, n. 1, p. 2787, 2023. Disponível em: <https://www.nature.com/articles/s41467-023-38347-2>. Acesso em: 01 de out. de 2023.

KUMAR, Gaurav *et al.* Artificial neural network-based modelling of optimized experimental study of xylanase production by *Penicillium citrinum* xym2. **Proceedings of the Institution of Mechanical Engineers, Part E: Journal of Process Mechanical Engineering**, [s. l.], v. 236, n. 4, p. 1340–1348, 2022. Disponível em: <http://journals.sagepub.com/doi/10.1177/09544089211064153>. Acesso em: 01 de out. de 2023.

KUMAR, L Ashok *et al.* Deep learning based assistive technology on audio visual speech recognition for hearing impaired. **International Journal of Cognitive Computing in Engineering**, [s. l.], v. 3, p. 24–30, 2022. Disponível em: <https://linkinghub.elsevier.com/retrieve/pii/S2666307422000031>. Acesso em: 01 de out. de 2023.

KUMAR, Chetan; CHOUDHARY, Alok. A top-down approach to classify enzyme functional classes and sub-classes using random forest. **EURASIP Journal on Bioinformatics and Systems Biology**, [s. l.], v. 2012, n. 1, p. 1, 2012. Disponível em: <https://bsb-urasipjournals.springeropen.com/articles/10.1186/1687-4153-2012-1>. Acesso em: 01 de out. de 2023.

KUMAR, Chetan; LI, Gang; CHOUDHARY, Alok. Enzyme Function Classification Using Protein Sequence Features and Random Forest. *In:* , 2009. **2009 3rd International Conference on Bioinformatics and Biomedical Engineering**. [S. l.]: IEEE, 2009. p. 1–4. Disponível em: <http://ieeexplore.ieee.org/document/5162790/>. Acesso em: 20 de ago. de 2023.

KUMARI, Madhulata; SUBBARAO, Naidu. Deep learning model for virtual screening of novel 3C-like protease enzyme inhibitors against SARS coronavirus diseases. **Computers in Biology and Medicine**, [s. l.], v. 132, p. 104317, 2021. Disponível em: <https://linkinghub.elsevier.com/retrieve/pii/S0010482521001116>. Acesso em: 25 fev. de 2023.

KUMI, Louis; JEONG, Jaewook. Optimization model for selecting optimal prefabricated column design considering environmental impacts and costs using genetic algorithm. **Journal of Cleaner Production**, [s. l.], v. 417, p. 137995, 2023. Disponível em: <https://linkinghub.elsevier.com/retrieve/pii/S0959652623021534>. Acesso em: 14 de out. de 2023.

KUNKA, Antonin *et al.* Advancing Enzyme's Stability and Catalytic Efficiency through Synergy of Force-Field Calculations, Evolutionary Analysis, and Machine Learning. **ACS Catalysis**, [s. l.], v. 13, n. 19, p. 12506–12518, 2023. Disponível em: <https://pubs.acs.org/doi/10.1021/acscatal.3c02575>. Acesso em: 30 de out. de 2023.

KUTYAUROIPO, Innocent; RUSHAMBWA, Munyaradzi; CHIWAZI, Lyndah. Artificial intelligence applications in the agrifood sectors. **Journal of Agriculture and Food Research**, [s. l.], v. 11, p. 100502, 2023. Disponível em: <https://linkinghub.elsevier.com/retrieve/pii/S2666154323000091>. Acesso em: 24 de fev. de 2023.

LEE, Ching-Hung *et al.* Understanding digital transformation in advanced manufacturing and engineering: A bibliometric analysis, topic modeling and research trend discovery. **Advanced Engineering Informatics**, [s. l.], v. 50, p. 101428, 2021. Disponível em: <https://linkinghub.elsevier.com/retrieve/pii/S1474034621001804>. Acesso em: 13 de nov. de 2023.

LEVIN, Itai *et al.* Merging enzymatic and synthetic chemistry with computational synthesis planning. **Nature Communications**, [s. l.], v. 13, n. 1, p. 7747, 2022. Disponível em: <https://www.nature.com/articles/s41467-022-35422-y>. Acesso em: 20 de ago. de 2023.

LEVY, Joshua *et al.* Artificial Intelligence, Bioinformatics, and Pathology. **Advances in Molecular Pathology**, [s. l.], 2023. Disponível em: <https://linkinghub.elsevier.com/retrieve/pii/S2589408023000017>. Acesso em: 16 de abr. de 2023.

LI, Lin *et al.* A bibliometric and visual analysis of global geo-ontology research. **Computers & Geosciences**, [s. l.], v. 99, p. 1–8, 2017. Disponível em: <https://linkinghub.elsevier.com/retrieve/pii/S0098300416305659>. Acesso em: 25 de fev. de 2023.

LI, Yaping *et al.* An Artificial Neural Network Assisted Optimization System for Analog Design Space Exploration. **IEEE Transactions on Computer-Aided Design of Integrated Circuits and Systems**, [s. l.], v. 39, n. 10, p. 2640–2653, 2020. Disponível em: <https://ieeexplore.ieee.org/document/8937720/>. Acesso em: 17 de abr. de 2023.

LI, Yu *et al.* DEEPRe: sequence-based enzyme EC number prediction by deep learning. **Bioinformatics**, [s. l.], v. 34, n. 5, p. 760–769, 2018. Disponível em: <https://academic.oup.com/bioinformatics/article/34/5/760/4562505>. Acesso em: 16 de abr. de 2023.

LI, Guangyue *et al.* Machine Learning Enables Selection of Epistatic Enzyme Mutants for Stability Against Unfolding and Detrimental Aggregation. **ChemBioChem**, [s. l.], v. 22, n. 5, p. 904–914, 2021. Disponível em: <https://chemistry-europe.onlinelibrary.wiley.com/doi/10.1002/cbic.202000612>. Acesso em: 02 de nov. de 2023.

LI, Guangyue; DONG, Yijie; REETZ, Manfred T. Can Machine Learning Revolutionize Directed Evolution of Selective Enzymes?. **Advanced Synthesis & Catalysis**, [s. l.], p. adsc.201900149, 2019a. Disponível em: <https://onlinelibrary.wiley.com/doi/10.1002/adsc.201900149>. Acesso em: 25 de fev. de 2023.

LIU, Yan *et al.* Enhancing the thermal stability of ketoreductase ChKRED12 using the FireProt web server. **Process Biochemistry**, [s. l.], v. 101, p. 207–212, 2021. Disponível em: <https://linkinghub.elsevier.com/retrieve/pii/S135951132031730X>. Acesso em: 30 de out. de 2023.

LUDERMIR, Teresa Bernarda. Inteligência Artificial e Aprendizado de Máquina: estado atual e tendências. **Estudos Avançados**, [s. l.], v. 35, n. 101, p. 85–94, 2021. Disponível em: http://www.scielo.br/scielo.php?script=sci_arttext&pid=S0103-40142021000100085&tlng=pt. Acesso em: 13 de nov. de 2023.

LUO, Jiefeng *et al.* Emerging Role of Artificial Intelligence in Diagnosis, Classification and Clinical Management of Glioma. **Seminars in Cancer Biology**, [s. l.], 2023. Disponível em: <https://linkinghub.elsevier.com/retrieve/pii/S1044579X23000457>. Acesso em: 13 de mar. de 2023.

LUO, Jingyang *et al.* Prediction of biological nutrients removal in full-scale wastewater treatment plants using H2O automated machine learning and back propagation artificial neural network model: Optimization and comparison. **Bioresource Technology**, [s. l.], p. 129842, 2023. Disponível em: <https://linkinghub.elsevier.com/retrieve/pii/S0960852423012701>. Acesso em: 13 de out. de 2023.

MAENPUEN, Somchart *et al.* Creating Flavin Reductase Variants with Thermostable and Solvent-Tolerant Properties by Rational-Design Engineering. **ChemBioChem**, [s. l.], v. 21, n. 10, p. 1481–1491, 2020. Disponível em: <https://chemistry-europe.onlinelibrary.wiley.com/doi/10.1002/cbic.201900737>. Acesso em: 30 de out. de 2023.

MAHDAVI-MEYMAND, Amin; SULISZ, Wojciech. Application of nested artificial neural network for the prediction of significant wave height. **Renewable Energy**, [s. l.], v. 209, p. 157–168, 2023. Disponível em: <https://linkinghub.elsevier.com/retrieve/pii/S0960148123004214>. Acesso em: 17 de abr. de 2023.

MAINALI, Shraddha; PARK, Soojin. Artificial Intelligence and Big Data Science in Neurocritical Care. **Critical Care Clinics**, [s. l.], v. 39, n. 1, p. 235–242, 2023. Disponível em: <https://linkinghub.elsevier.com/retrieve/pii/S0749070422000586>. Acesso em: 13 de mar. de 2023.

MALEKI, Nasim *et al.* A novel enzyme based biosensor for catechol detection in water samples using artificial neural network. **Biochemical Engineering Journal**, [s. l.], v. 128, p. 1–11, 2017. Disponível em: <https://linkinghub.elsevier.com/retrieve/pii/S1369703X17302255>. Acesso em: 15 de nov. de 2023.

MANIMARAN, A. *et al.* Application Areas, Benefits, and Research Challenges of Converging Blockchain and Machine Learning Techniques. *In: [S. l.: s. n.]*, 2024. p. 1–16. Disponível em: https://link.springer.com/10.1007/978-3-031-35751-0_1. Acesso em: 01 de nov. de 2023.

MAZURENKO, Stanislav; PROKOP, Zbynek; DAMBORSKY, Jiri. Machine Learning in Enzyme Engineering. **ACS Catalysis**, [s. l.], v. 10, n. 2, p. 1210–1223, 2020. Disponível em: <https://pubs.acs.org/doi/10.1021/acscatal.9b04321>. Acesso em: 07 de ago. de 2023.

MEI, Miao; TAN, Xu. Current Strategies of Antiviral Drug Discovery for COVID-19. **Frontiers in Molecular Biosciences**, [s. l.], v. 8, 2021. Disponível em: <https://www.frontiersin.org/articles/10.3389/fmolb.2021.671263/full>. Acesso em: 14 de nov. de 2023.

MEMON, Safyan Aman; AAMIR KHAN, Kinaan; NAVEED, Hammad. Enzyme Function Prediction using Deep Learning. **Biophysical Journal**, [s. l.], v. 118, n. 3, p. 533a, 2020. Disponível em: <https://linkinghub.elsevier.com/retrieve/pii/S0006349519338597>. Acesso em: 25 de fev. de 2023.

MENG, Qiaozhen; GUO, Fei; TANG, Jijun. Improved structure-related prediction for insufficient homologous proteins using MSA enhancement and pre-trained language model. **Briefings in Bioinformatics**, [s. l.], v. 24, n. 4, 2023. Disponível em: <https://academic.oup.com/bib/article/doi/10.1093/bib/bbad217/7198547>. Acesso em: 30 de out. de 2023.

MICHAEL HALL, C. Publish and perish? Bibliometric analysis, journal ranking and the assessment of research quality in tourism. **Tourism Management**, [s. l.], v. 32, n. 1, p. 16–27, 2011. Disponível em: <https://linkinghub.elsevier.com/retrieve/pii/S0261517710001469>. Acesso em: 31 de jul. de 2023.

MING, Yongfan *et al.* A review of enzyme design in catalytic stability by artificial intelligence. **Briefings in Bioinformatics**, [s. l.], v. 24, n. 3, 2023a. Disponível em: <https://academic.oup.com/bib/article/doi/10.1093/bib/bbad065/7086816>. Acesso em: 30 de out. de 2023.

MITCHELL, Melanie. **An Introduction to Genetic Algorithms**. [S. l.]: The MIT Press, 1998. Disponível em: <https://mitpress.mit.edu/9780262631853/an-introduction-to-genetic-algorithms/>. Acesso em: 14 de out. de 2023.

MOAYEDI, Hossein *et al.* A systematic review and meta-analysis of artificial neural network application in geotechnical engineering: theory and applications. **Neural Computing and Applications**, [s. l.], v. 32, n. 2, p. 495–518, 2020. Disponível em: <http://link.springer.com/10.1007/s00521-019-04109-9>. Acesso em: 17 de abr. de 2023.

MOHEBBI, M. *et al.* Modeling and optimization of viscosity in enzyme-modified cheese by fuzzy logic and genetic algorithm. **Computers and Electronics in Agriculture**, [s. l.], v. 62, n. 2, p. 260–265, 2008. Disponível em: <https://linkinghub.elsevier.com/retrieve/pii/S016816990800015X>. Acesso em: 15 de nov. de 2023.

MORA, Luca; BOLICI, Roberto; DEAKIN, Mark. The First Two Decades of Smart-City Research: A Bibliometric Analysis. **Journal of Urban Technology**, [s. l.], v. 24, n. 1, p. 3–27, 2017. Disponível em: <https://www.tandfonline.com/doi/full/10.1080/10630732.2017.1285123>. Acesso em: 04 de mai. de 2023.

MOU, Zhongyu *et al.* Machine learning-based prediction of enzyme substrate scope: Application to bacterial nitrilases. **Proteins: Structure, Function, and Bioinformatics**, [s. l.], v. 89, n. 3, p. 336–347, 2021. Disponível em: <https://onlinelibrary.wiley.com/doi/10.1002/prot.26019>. Acesso em: 01 de out. de 2023.

MULFARI, Davide *et al.* Machine learning assistive application for users with speech disorders. **Applied Soft Computing**, [s. l.], v. 103, p. 107147, 2021. Disponível em: <https://linkinghub.elsevier.com/retrieve/pii/S1568494621000703>. Acesso em: 13 de mar. de 2023.

MUSIL, Milos *et al.* FireProtASR: A Web Server for Fully Automated Ancestral Sequence Reconstruction. **Briefings in Bioinformatics**, [s. l.], v. 22, n. 4, 2021. Disponível em: <https://academic.oup.com/bib/article/doi/10.1093/bib/bbaa337/6042664>. Acesso em: 30 de out. de 2023.

MUTHUKRISHNAN, Nikesh *et al.* Brief History of Artificial Intelligence. **Neuroimaging Clinics of North America**, [s. l.], v. 30, n. 4, p. 393–399, 2020. Disponível em: <https://linkinghub.elsevier.com/retrieve/pii/S105251492030054X>. Acesso em: 13 de mar. de 2023.

MYSTRIDIS, George A. *et al.* Artificial Intelligence/Machine Learning and Mechanistic Modeling Approaches as Translational Tools to Advance Personalized Medicine Decisions. **Advances in Molecular Pathology**, [s. l.], v. 5, n. 1, p. 131–139, 2022. Disponível em: <https://linkinghub.elsevier.com/retrieve/pii/S2589408022000072>. Acesso em: 24 de fev. de 2023.

NALLAPAREDDY, Mohan Vamsi; DWIVEDULA, Rohit. ABLE: Attention based learning for enzyme classification. **Computational Biology and Chemistry**, [s. l.], v. 94, p. 107558, 2021. Disponível em: <https://linkinghub.elsevier.com/retrieve/pii/S1476927121001250>. Acesso em: 13 de nov. de 2023.

NANDAKUMAR, Rohit; DINU, Valentin. Developing a machine learning model to identify protein–protein interaction hotspots to facilitate drug discovery. **PeerJ**, [s. l.], v. 8, p. e10381, 2020. Disponível em: <https://peerj.com/articles/10381>. Acesso em: 02 de nov. de 2023.

NIEL, Olivier; BASTARD, Paul. Artificial Intelligence in Nephrology: Core Concepts, Clinical Applications, and Perspectives. **American Journal of Kidney Diseases**, [s. l.], v. 74, n. 6, p. 803–810, 2019. Disponível em: <https://linkinghub.elsevier.com/retrieve/pii/S027263861930842X>. Acesso em: 13 de mar. de 2023.

NIU, Bing *et al.* Predicting network of drug–enzyme interaction based on machine learning method. **Biochimica et Biophysica Acta (BBA) - Proteins and Proteomics**, [s. l.], v. 1844, n. 1, p. 214–223, 2014. Disponível em: <https://linkinghub.elsevier.com/retrieve/pii/S157096391300280X>. Acesso em: 06 de ago. de 2023.

NOUROUZI, Zohreh; CHAMANI, Atefeh. Characterization of ambient carbon monoxide and PM 2.5 effects on fetus development, liver enzymes and TSH in Isfahan City, central Iran. **Environmental Pollution**, [s. l.], v. 291, p. 118238, 2021. Disponível em: <https://linkinghub.elsevier.com/retrieve/pii/S0269749121018200>. Acesso em: 01 de out. de 2023.

NUNES, Yale Luck *et al.* Chemical and physical Chitosan modification for designing enzymatic industrial biocatalysts: How to choose the best strategy?. **International Journal of Biological Macromolecules**, [s. l.], v. 181, p. 1124–1170, 2021. Disponível em: <https://linkinghub.elsevier.com/retrieve/pii/S0141813021007595>. Acesso em: 25 de fev. de 2023.

OKAFOR, Christian Emeka *et al.* Advances in machine learning-aided design of reinforced polymer composite and hybrid material systems. **Hybrid Advances**, [s. l.], v. 2, p. 100026, 2023. Disponível em: <https://linkinghub.elsevier.com/retrieve/pii/S2773207X2300009X>. Acesso em: 13 de mar. de 2023.

OKUYUCU, Hasan; KURT, Adem; ARCAKLIOGLU, Erol. Artificial neural network application to the friction stir welding of aluminum plates. **Materials & Design**, [s. l.], v. 28, n. 1, p. 78–84, 2007. Disponível em: <https://linkinghub.elsevier.com/retrieve/pii/S0261306905001597>. Acesso em: 17 de abr. de 2023.

OLABI, A.G. *et al.* Application of artificial intelligence for prediction, optimization, and control of thermal energy storage systems. **Thermal Science and Engineering Progress**, [s. l.], p. 101730, 2023. Disponível em: <https://linkinghub.elsevier.com/retrieve/pii/S2451904923000835>. Acesso em: 24 de fev. de 2023.

ORTIZ, Claudia *et al.* Novozym 435: the “perfect” lipase immobilized biocatalyst? **Catalysis Science & Technology**, [s. l.], v. 9, n. 10, p. 2380–2420, 2019. Disponível em: <http://xlink.rsc.org/?DOI=C9CY00415G>. Acesso em: 25 de fev. de 2023.

OSTAD-ALI-ASKARI, Kaveh; SHAYANNEJAD, Mohammad; GHORBANIZADEH-KHARAZI, Hossein. Artificial neural network for modeling nitrate pollution of groundwater in marginal area of Zayandeh-rood River, Isfahan, Iran. **KSCE Journal of Civil Engineering**, [s. l.], v. 21, n. 1, p. 134–140, 2017. Disponível em: <http://link.springer.com/10.1007/s12205-016-0572-8>. Acesso em: 17 de abr. de 2023.

OUSAADI, Mouna Imene *et al.* Valorization and optimization of agro-industrial orange waste for the production of enzyme by halophilic *Streptomyces* sp. **Environmental Research**, [s. l.], v. 201, p. 111494, 2021. Disponível em: <https://linkinghub.elsevier.com/retrieve/pii/S001393512100788X>. Acesso em: 01 de out. de 2023.

OWOLABI, Titilayo Abimbola; SAJJAD, Muhammad. A global outlook on multi-hazard risk analysis: A systematic and scientometric review. **International Journal of Disaster Risk Reduction**, [s. l.], p. 103727, 2023. Disponível em: <https://linkinghub.elsevier.com/retrieve/pii/S2212420923002078>. Acesso em: 14 de mai. de 2023.

PACAL, Ishak *et al.* A comprehensive review of deep learning in colon cancer. **Computers in Biology and Medicine**, [s. l.], v. 126, p. 104003, 2020. Disponível em: <https://linkinghub.elsevier.com/retrieve/pii/S0010482520303346>. Acesso em: 13 de mar. de 2023.

PAHIKKALA, T. *et al.* Toward more realistic drug-target interaction predictions. **Briefings in Bioinformatics**, [s. l.], v. 16, n. 2, p. 325–337, 2015. Disponível em: <https://academic.oup.com/bib/article-lookup/doi/10.1093/bib/bbu010>. Acesso em: 27 de mai. de 2023.

PALADINO, Antonella *et al.* Protein design: from computer models to artificial intelligence. **WIREs Computational Molecular Science**, [s. l.], v. 7, n. 5, 2017. Disponível em: <https://wires.onlinelibrary.wiley.com/doi/10.1002/wcms.1318>. Acesso em: 05 de out. de 2023.

PALLAVI, Pogaku *et al.* Optimization of lipase production by *Staphylococcus* sp. Lp12. **African Journal of Biotechnology**, [s. l.], v. 9, n. 6, p. 882–886, 2010. Disponível em: <http://academicjournals.org/journal/AJB/article-abstract/88F587D34902>. Acesso em: 02 de nov. de 2023.

PARIA, Prasenjit; TASSANAKAJON, Anchalee. Identification of Potential Druggable Targets and Structure-Based Virtual Screening for Drug-like Molecules against the Shrimp Pathogen *Enterocytozoon hepatopenaei*. **International Journal of Molecular Sciences**, [s. l.], v. 24, n. 2, p. 1412, 2023. Disponível em: <https://www.mdpi.com/1422-0067/24/2/1412>. Acesso em: 04 de out. de 2023.

PATENTSCOPE. **World Intellectual Property Organization (WIPO)**. [s. l.], 2023. Disponível em: <https://patentscope.wipo.int/search/pt/search.jsf>. Acesso em: 19 out. 2023.

PAULETTO, Paola S.; DOTTO, Guilherme L.; SALAU, Nina P.G. Optimal artificial neural network design for simultaneous modeling of multicomponent adsorption. **Journal of Molecular Liquids**, [s. l.], v. 320, p. 114418, 2020. Disponível em: <https://linkinghub.elsevier.com/retrieve/pii/S0167732220353058>. Acesso em: 17 de abr. de 2023.

PENG, Yujia *et al.* The Tong Test: Evaluating Artificial General Intelligence Through Dynamic Embodied Physical and Social Interactions. **Engineering**, [s. l.], 2023. Disponível em: <https://linkinghub.elsevier.com/retrieve/pii/S209580992300293X>. Acesso em: 13 de nov. de 2023.

PLONKA, Wojciech *et al.* CYLebrity: Machine learning models for the prediction of inhibitors of cytochrome P450 enzymes. **Bioorganic & Medicinal Chemistry**, [s. l.], v. 46, p. 116388, 2021. Disponível em: <https://linkinghub.elsevier.com/retrieve/pii/S0968089621003965>. Acesso em: 25 de fev. de 2023.

POREBSKI, Benjamin T.; BUCKLE, Ashley M. Consensus protein design. **Protein Engineering Design and Selection**, [s. l.], v. 29, n. 7, p. 245–251, 2016. Disponível em: <https://academic.oup.com/peds/article-lookup/doi/10.1093/protein/gzw015>. Acesso em: 30 de out. de 2023.

PRADHAN, Debasmita; PADHY, Sudarsan; SAHOO, Biswajit. Enzyme classification using multiclass support vector machine and feature subset selection. **Computational Biology and Chemistry**, [s. l.], v. 70, p. 211–219, 2017. Disponível em: <https://linkinghub.elsevier.com/retrieve/pii/S1476927117303298>. Acesso em: 15 de nov. de 2023.

QU, Ge *et al.* The Crucial Role of Methodology Development in Directed Evolution of Selective Enzymes. **Angewandte Chemie International Edition**, [s. l.], v. 59, n. 32, p. 13204–13231, 2020. Disponível em: <https://onlinelibrary.wiley.com/doi/10.1002/anie.201901491>. Acesso em: 27 de mai. de 2023.

RALBOVSKY, Nicole M.; SMITH, Joseph P. Machine Learning and Chemical Imaging to Elucidate Enzyme Immobilization for Biocatalysis. **Analytical Chemistry**, [s. l.], v. 93, n. 35, p. 11973–11981, 2021. Disponível em: <https://pubs.acs.org/doi/10.1021/acs.analchem.1c01909>. Acesso em: 30 de out. de 2023.

RANJBARI, Meisam *et al.* Biomass and organic waste potentials towards implementing circular bioeconomy platforms: A systematic bibliometric analysis. **Fuel**, [s. l.], v. 318, p. 123585, 2022. Disponível em: <https://linkinghub.elsevier.com/retrieve/pii/S0016236122004501>. Acesso em: 12 de mar. de 2023.

REJEB, Abderahman *et al.* The big picture on Instagram research: Insights from a bibliometric analysis. **Telematics and Informatics**, [s. l.], v. 73, p. 101876, 2022. Disponível em: <https://linkinghub.elsevier.com/retrieve/pii/S0736585322001095>. Acesso em: 11 de nov. de 2023.

RIBELLES, Nuria *et al.* Machine learning and natural language processing (NLP) approach to predict early progression to first-line treatment in real-world hormone receptor-positive (HR+)/HER2-negative advanced breast cancer patients. **European Journal of Cancer**, [s. l.], v. 144, p. 224–231, 2021. Disponível em: <https://linkinghub.elsevier.com/retrieve/pii/S0959804920313678>. Acesso em: 13 de mar. de 2023.

RIUL JR., Antonio *et al.* Recent advances in electronic tongues. **The Analyst**, [s. l.], v. 135, n. 10, p. 2481, 2010. Disponível em: <http://xlink.rsc.org/?DOI=c0an00292e>. Acesso em: 27 de mai. de 2023.

ROBINSON, Serina L.; PIEL, Jörn; SUNAGAWA, Shinichi. A roadmap for metagenomic enzyme discovery. **Natural Product Reports**, [s. l.], v. 38, n. 11, p. 1994–2023, 2021. Disponível em: <http://xlink.rsc.org/?DOI=D1NP00006C>. Acesso em: 01 de nov. de 2023.

ROCHE, Rahmatullah *et al.* E(3) equivariant graph neural networks for robust and accurate protein-protein interaction site prediction. **PLOS Computational Biology**, [s. l.], v. 19, n. 8, p. e1011435, 2023. Disponível em: <https://dx.plos.org/10.1371/journal.pcbi.1011435>. Acesso em: 05 de out. de 2023.

RODRIGUES, Antonio F.S. *et al.* A scientometric analysis of research progress and trends in the design of laccase biocatalysts for the decolorization of synthetic dyes. **Process Biochemistry**, [s. l.], v. 126, p. 272–291, 2023. Disponível em: <https://linkinghub.elsevier.com/retrieve/pii/S135951132300020X>. Acesso em: 11 de mar. de 2023.

RODRÍGUEZ-SOLER, Rocío; URIBE-TORIL, Juan; DE PABLO VALENCIANO, Jaime. Worldwide trends in the scientific production on rural depopulation, a bibliometric analysis using bibliometrix R-tool. **Land Use Policy**, [s. l.], v. 97, p. 104787, 2020. Disponível em: <https://linkinghub.elsevier.com/retrieve/pii/S0264837719314450>. Acesso em: 14 de mai. de 2023.

ROY, Kunal; ROY, Partha Pratim. Exploring QSAR and QAAR for inhibitors of cytochrome P450 2A6 and 2A5 enzymes using GFA and G/PLS techniques. **European Journal of Medicinal Chemistry**, [s. l.], v. 44, n. 5, p. 1941–1951, 2009. Disponível em: <https://linkinghub.elsevier.com/retrieve/pii/S0223523408005801>. Acesso em: 01 de out. de 2023.

RYU, Gahyeon *et al.* Deep learning for metabolic pathway design. **Metabolic Engineering**, [s. l.], v. 80, p. 130–141, 2023. Disponível em: <https://linkinghub.elsevier.com/retrieve/pii/S1096717623001386>. Acesso em: 01 de nov. de 2023.

RYU, Jae Yong; KIM, Hyun Uk; LEE, Sang Yup. Deep learning enables high-quality and high-throughput prediction of enzyme commission numbers. **Proceedings of the National Academy of Sciences**, [s. l.], v. 116, n. 28, p. 13996–14001, 2019. Disponível em: <https://pnas.org/doi/full/10.1073/pnas.1821905116>. Acesso em: 01 de out. de 2023.

SAFDAR, Rizwan *et al.* A bibliometric analysis for estimating the global research trends related to applications of ionic liquids in drug delivery. **Journal of Molecular Liquids**, [s. l.], v. 381, p. 121778, 2023. Disponível em: <https://linkinghub.elsevier.com/retrieve/pii/S0167732223005810>. Acesso em: 16 de abr. de 2023.

SALEHI, Hadi; BURGUEÑO, Rigoberto. Emerging artificial intelligence methods in structural engineering. **Engineering Structures**, [s. l.], v. 171, p. 170–189, 2018. Disponível em: <https://linkinghub.elsevier.com/retrieve/pii/S0141029617335526>. Acesso em: 13 de mar. de 2023.

SALES, Misael B. *et al.* Sustainable Feedstocks and Challenges in Biodiesel Production: An Advanced Bibliometric Analysis. **Bioengineering**, [s. l.], v. 9, n. 10, p. 539, 2022. Disponível em: <https://www.mdpi.com/2306-5354/9/10/539>. Acesso em: 11 de mar. de 2023.

SALES, Misael Bessa *et al.* Trends and Opportunities in Enzyme Biosensors Coupled to Metal-Organic Frameworks (MOFs): An Advanced Bibliometric Analysis. **Electrochem**, [s. l.], v. 4, n. 2, p. 181–211, 2023. Disponível em: <https://www.mdpi.com/2673-3293/4/2/14>. Acesso em: 16 de abr. de 2023.

SAMADI, Aryan; POUR, Azadeh Kermanshahi; JAMIESON, Rob. Development of remediation technologies for organic contaminants informed by QSAR/QSPR models. **Environmental Advances**, [s. l.], v. 5, p. 100112, 2021. Disponível em: <https://linkinghub.elsevier.com/retrieve/pii/S2666765721000831>. Acesso em: 18 de out. de 2023.

SAMPAIO, Pedro Sousa; FERNANDES, Pedro. Machine Learning: A Suitable Method for Biocatalysis. **Catalysts**, [s. l.], v. 13, n. 6, p. 961, 2023. Disponível em: <https://www.mdpi.com/2073-4344/13/6/961>. Acesso em: 05 de out. de 2023.

SARNO, Laura *et al.* Use of artificial intelligence in obstetrics: not quite ready for prime time. **American Journal of Obstetrics & Gynecology MFM**, [s. l.], v. 5, n. 2, p. 100792, 2023. Disponível em: <https://linkinghub.elsevier.com/retrieve/pii/S2589933322002221>. Acesso em: 13 de mar. de 2023.

SCHNEIDER, Susan. Superintelligent AI and the Postbiological Cosmos Approach¹. *In*: WHAT IS LIFE? ON EARTH AND BEYOND. [S. l.]: Cambridge University Press, 2017. p. 178–198. Disponível em: https://www.cambridge.org/core/product/identifiier/9781316809648%23CT-bp-10/type/book_part. Acesso em: 13 de nov. de 2023.

SCHRAMM, Simon; PIEPER, Matthias; VOGL, Stefanie. Orthogonal Procrustes Based Anomaly Detection and Error Prediction for Vehicle Bills of Materials. **SSRN Electronic Journal**, [s. l.], 2022. Disponível em: <https://www.ssrn.com/abstract=4120321>. Acesso em: 13 de nov. de 2023.

SCHRAMM, Simon; WEHNER, Christoph; SCHMID, Ute. Comprehensible Artificial Intelligence on Knowledge Graphs: A survey. **Journal of Web Semantics**, [s. l.], v. 79, p. 100806, 2023. Disponível em: <https://linkinghub.elsevier.com/retrieve/pii/S1570826823000355>. Acesso em: 13 de nov. de 2023.

SEQUEIRA, Ana Marta; LOUSA, Diana; ROCHA, Miguel. ProPythia: A Python package for protein classification based on machine and deep learning. **Neurocomputing**, [s. l.], v. 484, p. 172–182, 2022. Disponível em: <https://linkinghub.elsevier.com/retrieve/pii/S0925231221016568>. Acesso em: 14 de nov. de 2023.

SHAH, Harshil; CHAVDA, Vivek; SONIWALA, Moinuddin M. Applications of Bioinformatics Tools in Medicinal Biology and Biotechnology. *In*: BIOINFORMATICS TOOLS FOR PHARMACEUTICAL DRUG PRODUCT DEVELOPMENT. [S. l.]: Wiley, 2023. p. 95–116. Disponível em: <https://onlinelibrary.wiley.com/doi/10.1002/9781119865728.ch6>. Acesso em: 02 de nov. de 2023.

SHAHARE, Yogesh *et al.* A Comprehensive Analysis of Machine Learning-Based Assessment and Prediction of Soil Enzyme Activity. **Agriculture**, [s. l.], v. 13, n. 7, p. 1323, 2023. Disponível em: <https://www.mdpi.com/2077-0472/13/7/1323>. Acesso em: 13 de nov. de 2023.

SHAHID, Saman; MASOOD, Khalid; KHAN, Abdul Waheed. Prediction of impacts on liver enzymes from the exposure of low-dose medical radiations through artificial intelligence algorithms. **Revista da Associação Médica Brasileira**, [s. l.], v. 67, n. 2, p. 248–259, 2021. Disponível em: http://www.scielo.br/scielo.php?script=sci_arttext&pid=S0104-42302021000300248&tlng=en. Acesso em: 01 de out. de 2023.

SHARIF, Mahir M. *et al.* Enzyme vs. non-enzyme classification based on principal component analysis and AdaBoost classifier. *In:* , 2016. **2016 International Conference on Computing, Communication and Automation (ICCCA)**. [S. l.]: IEEE, 2016. p. 288–293. Disponível em: <http://ieeexplore.ieee.org/document/7813731/>. Acesso em: 01 de out. de 2023.

SHEIKH, Ahmad Y.; FANN, James I. Artificial Intelligence. **Thoracic Surgery Clinics**, [s. l.], v. 29, n. 3, p. 339–350, 2019. Disponível em: <https://linkinghub.elsevier.com/retrieve/pii/S1547412719300222>. Acesso em: 24 de fev. de 2023.

SHELDON, Roger A.; VAN PELT, Sander. Enzyme immobilisation in biocatalysis: why, what and how. **Chem. Soc. Rev.**, [s. l.], v. 42, n. 15, p. 6223–6235, 2013. Disponível em: <http://xlink.rsc.org/?DOI=C3CS60075K>. Acesso em: 25 de fev. de 2023.

SHOME, Samik *et al.* Impact investment for sustainable development: A bibliometric analysis. **International Review of Economics & Finance**, [s. l.], v. 84, p. 770–800, 2023. Disponível em: <https://linkinghub.elsevier.com/retrieve/pii/S105905602200315X>. Acesso em: 12 de mar. de 2023.

SHROFF, Raghav *et al.* Discovery of Novel Gain-of-Function Mutations Guided by Structure-Based Deep Learning. **ACS Synthetic Biology**, [s. l.], v. 9, n. 11, p. 2927–2935, 2020. Disponível em: <https://pubs.acs.org/doi/10.1021/acssynbio.0c00345>. Acesso em: 30 de out. de 2023.

SIEDHOFF, Niklas E.; SCHWANEBERG, Ulrich; DAVARI, Mehdi D. Machine learning-assisted enzyme engineering. *In:* [S. l.: s. n.], 2020. p. 281–315. Disponível em: <https://linkinghub.elsevier.com/retrieve/pii/S0076687920302275>. Acesso em: 07 de ago. de 2023.

SILVA, Allison *et al.* The Chemistry and Applications of Metal–Organic Frameworks (MOFs) as Industrial Enzyme Immobilization Systems. **Molecules**, [s. l.], v. 27, n. 14, p. 4529, 2022. Disponível em: <https://www.mdpi.com/1420-3049/27/14/4529>. Acesso em: 25 de fev. de 2023.

SINGH, Nitu *et al.* Revolutionizing enzyme engineering through artificial intelligence and machine learning. **Emerging Topics in Life Sciences**, [s. l.], v. 5, n. 1, p. 113–125, 2021. Disponível em: <https://portlandpress.com/emergtoplifesci/article/5/1/113/228244/Revolutionizing-enzyme-engineering-through>. Acesso em: 14 de out. de 2023.

SOUZA, José E. da S. *et al.* A Comprehensive Review on the Use of Metal–Organic Frameworks (MOFs) Coupled with Enzymes as Biosensors. **Electrochem**, [s. l.], v. 3, n. 1, p. 89–113, 2022. Disponível em: <https://www.mdpi.com/2673-3293/3/1/6>. Acesso em: 25 de fev. de 2023.

SRAPHET, Supajit; JAVADI, Bagher. Application of Hierarchical Clustering to Analyze Solvent-Accessible Surface Area Patterns in Amycolatopsis lipases. **Biology**, [s. l.], v. 11, n. 5, p. 652, 2022. Disponível em: <https://www.mdpi.com/2079-7737/11/5/652>. Acesso em: 30 de out. de 2023.

SRAPHET, Supajit; JAVADI, Bagher. Computational characterizations of GDP-mannose 4,6-dehydratase (NoeL) Rhizobial proteins. **Current Genetics**, [s. l.], v. 67, n. 5, p. 769–784, 2021. Disponível em: <https://link.springer.com/10.1007/s00294-021-01184-1>. Acesso em: 30 de out. de 2023.

STEINER, Kerstin; SCHWAB, Helmut. RECENT ADVANCES IN RATIONAL APPROACHES FOR ENZYME ENGINEERING. **Computational and Structural Biotechnology Journal**, [s. l.], v. 2, n. 3, p. e201209010, 2012. Disponível em: <https://linkinghub.elsevier.com/retrieve/pii/S200103701460088X>. Acesso em: 07 de ago. de 2023.

SYED, Umar; YONA, Golan. Enzyme Function Prediction with Interpretable Models. *In*: [S. l.: s. n.], 2009. p. 373–420. Disponível em: http://link.springer.com/10.1007/978-1-59745-243-4_17. Acesso em: 01 de out. de 2023.

SZALENIEC, Maciej. Prediction of enzyme activity with neural network models based on electronic and geometrical features of substrates. **Pharmacological Reports**, [s. l.], v. 64, n. 4, p. 761–781, 2012. Disponível em: <https://linkinghub.elsevier.com/retrieve/pii/S1734114012708733>. Acesso em: 01 de out. de 2023.

TABAAZA, Grace Amabel *et al.* Application of machine learning models to predict cytotoxicity of ionic liquids using VolSurf principal properties. **Computational Toxicology**, [s. l.], v. 26, p. 100266, 2023. Disponível em: <https://linkinghub.elsevier.com/retrieve/pii/S2468111323000075>. Acesso em: 02 de nov. de 2023.

TAMALA, Justine Klaire *et al.* A bibliometric analysis of sustainable oil and gas production research using VOSviewer. **Cleaner Engineering and Technology**, [s. l.], v. 7, p. 100437, 2022. Disponível em: <https://linkinghub.elsevier.com/retrieve/pii/S2666790822000428>. Acesso em: 13 de mar. de 2023.

TANG, Yuk Ming; KUO, Wei Ting; LEE, C.K.M. Real-time Mixed Reality (MR) and Artificial Intelligence (AI) object recognition integration for digital twin in Industry 4.0. **Internet of Things**, [s. l.], v. 23, p. 100753, 2023. Disponível em: <https://linkinghub.elsevier.com/retrieve/pii/S2542660523000768>. Acesso em: 15 de nov. de 2023.

TAO, Fei; QI, Qinglin. Make more digital twins. **Nature**, [s. l.], v. 573, n. 7775, p. 490–491, 2019. Disponível em: <https://www.nature.com/articles/d41586-019-02849-1>. Acesso em: 02 de nov. de 2023.

TEJERA, Eduardo *et al.* Drugs Repurposing Using QSAR, Docking and Molecular Dynamics for Possible Inhibitors of the SARS-CoV-2 Mpro Protease. **Molecules**, [s. l.], v. 25, n. 21, p. 5172, 2020. Disponível em: <https://www.mdpi.com/1420-3049/25/21/5172>. Acesso em: 02 de nov. de 2023.

TERLOUW, Barbara R *et al.* MIBiG 3.0: a community-driven effort to annotate experimentally validated biosynthetic gene clusters. **Nucleic Acids Research**, [s. l.], v. 51, n. D1, p. D603–D610, 2023. Disponível em: <https://academic.oup.com/nar/article/51/D1/D603/6833236>. Acesso em: 01 de nov. de 2023.

TIAN, Yingjie *et al.* When Liver Disease Diagnosis Encounters Deep Learning: Analysis, Challenges, and Prospects. **iLIVER**, [s. l.], 2023. Disponível em: <https://linkinghub.elsevier.com/retrieve/pii/S2772947823000087>. Acesso em: 13 de mar. de 2023.

TUAN HOANG, Anh *et al.* A review on application of artificial neural network (ANN) for performance and emission characteristics of diesel engine fueled with biodiesel-based fuels. **Sustainable Energy Technologies and Assessments**, [s. l.], v. 47, p. 101416, 2021. Disponível em: <https://linkinghub.elsevier.com/retrieve/pii/S2213138821004264>. Acesso em: 17 de abr. de 2023.

TUNG, Cheng-Che *et al.* Optimization of mechanical properties of bio-inspired Voronoi structures by genetic algorithm. **Journal of Materials Research and Technology**, [s. l.], v. 26, p. 3813–3829, 2023. Disponível em: <https://linkinghub.elsevier.com/retrieve/pii/S2238785423020239>. Acesso em: 14 de out. de 2023.

UPADHYAY, Vikas; BOORLA, Veda Sheersh; MARANAS, Costas D. Rank-ordering of known enzymes as starting points for re-engineering novel substrate activity using a convolutional neural network. **Metabolic Engineering**, [s. l.], v. 78, p. 171–182, 2023. Disponível em: <https://linkinghub.elsevier.com/retrieve/pii/S1096717623000836>. Acesso em: 20 de ago. de 2023.

VALDÉS-TRESANCO, Mario S. *et al.* Structural Modeling of Nanobodies: A Benchmark of State-of-the-Art Artificial Intelligence Programs. **Molecules**, [s. l.], v. 28, n. 10, p. 3991, 2023. Disponível em: <https://www.mdpi.com/1420-3049/28/10/3991>. Acesso em: 30 de out. de 2023.

VALIPOUR, Mohammad; BANIHABIB, Mohammad Ebrahim; BEHBAHANI, Seyyed Mahmood Reza. Comparison of the ARMA, ARIMA, and the autoregressive artificial neural network models in forecasting the monthly inflow of Dez dam reservoir. **Journal of Hydrology**, [s. l.], v. 476, p. 433–441, 2013. Disponível em: <https://linkinghub.elsevier.com/retrieve/pii/S002216941200981X>. Acesso em: 17 de abr. de 2023.

VAN DER WEG, Karel J; GOHLKE, Holger. TopEnzyme: a framework and database for structural coverage of the functional enzyme space. **Bioinformatics**, [s. l.], v. 39, n. 3, 2023. Disponível em: <https://academic.oup.com/bioinformatics/article/doi/10.1093/bioinformatics/btad116/7072462>. Acesso em: 30 de out. de 2023.

VAN LAARHOVEN, Twan; MARCHIORI, Elena. Predicting Drug-Target Interactions for New Drug Compounds Using a Weighted Nearest Neighbor Profile. **PLoS ONE**, [s. l.], v. 8, n. 6, p. e66952, 2013. Disponível em: <https://dx.plos.org/10.1371/journal.pone.0066952>. Acesso em: 27 de mai. de 2023.

VANELLA, Rosario *et al.* High-throughput screening, next generation sequencing and machine learning: advanced methods in enzyme engineering. **Chemical Communications**, [s. l.], v. 58, n. 15, p. 2455–2467, 2022. Disponível em: <http://xlink.rsc.org/?DOI=D1CC04635G>. Acesso em: 14 de ago. de 2023.

VELASCO-LOZANO, Susana; ROCHA-MARTIN, Javier; SANTOS, José C. S. dos. Editorial: Designing Carrier-Free Immobilized Enzymes for Biocatalysis. **Frontiers in Bioengineering and Biotechnology**, [s. l.], v. 10, 2022. Disponível em: <https://www.frontiersin.org/articles/10.3389/fbioe.2022.924743/full>. Acesso em: 25 de fev. de 2023.

VOLK, Michael Jeffrey *et al.* Biosystems Design by Machine Learning. **ACS Synthetic Biology**, [s. l.], v. 9, n. 7, p. 1514–1533, 2020. Disponível em: <https://pubs.acs.org/doi/10.1021/acssynbio.0c00129>. Acesso em: 14 de nov. de 2023.

VORA, Lalitkumar K. *et al.* Artificial Intelligence in Pharmaceutical Technology and Drug Delivery Design. **Pharmaceutics**, [s. l.], v. 15, n. 7, p. 1916, 2023. Disponível em: <https://www.mdpi.com/1999-4923/15/7/1916>. Acesso em: 02 de nov. de 2023.

WANG, Jiamin *et al.* A term function-aware keyword citation network method for science mapping analysis. **Information Processing & Management**, [s. l.], v. 60, n. 4, p. 103405, 2023. Disponível em: <https://linkinghub.elsevier.com/retrieve/pii/S0306457323001425>. Acesso em: 18 de out. de 2023.

WANG, Xinglong *et al.* AI-assisted food enzymes design and engineering: a critical review. **Systems Microbiology and Biomanufacturing**, [s. l.], v. 3, n. 1, p. 75–87, 2023. Disponível em: <https://link.springer.com/10.1007/s43393-022-00138-z>. Acesso em: 01 de out. de 2023.

WANG, Youzhao *et al.* Electric heating promotes sludge composting process: Optimization of heating method through machine learning algorithms. **Bioresource Technology**, [s. l.], v. 382, p. 129177, 2023. Disponível em: <https://linkinghub.elsevier.com/retrieve/pii/S096085242300603X>. Acesso em: 02 de out. de 2023.

WANG, Yali *et al.* Enzyme engineering: from artificial design to artificial intelligence. **Huagong Xuebao/CIESC Journal**, [s. l.], v. 72, n. 7, p. 3590–3600, 2021. Acesso em: 13 de out. de 2023.

WANG, Zhen-Hua *et al.* Prediction of bending force in the hot strip rolling process using artificial neural network and genetic algorithm (ANN-GA). **The International Journal of Advanced Manufacturing Technology**, [s. l.], v. 93, n. 9–12, p. 3325–3338, 2017. Disponível em: <http://link.springer.com/10.1007/s00170-017-0711-5>. Acesso em: 17 de abr. de 2023.

WATANABE, Naoki *et al.* Comprehensive Machine Learning Prediction of Extensive Enzymatic Reactions. **The Journal of Physical Chemistry B**, [s. l.], v. 126, n. 36, p. 6762–6770, 2022. Disponível em: <https://pubs.acs.org/doi/10.1021/acs.jpcc.2c03287>. Acesso em: 01 de nov. de 2023.

WATANABE, Naoki *et al.* Exploration and Evaluation of Machine Learning-Based Models for Predicting Enzymatic Reactions. **Journal of Chemical Information and Modeling**, [s. l.], v. 60, n. 3, p. 1833–1843, 2020. Disponível em: <https://pubs.acs.org/doi/10.1021/acs.jcim.9b00877>. Acesso em: 01 de nov. de 2023.

WATSON, Joseph L. *et al.* De novo design of protein structure and function with RFDiffusion. **Nature**, [s. l.], v. 620, n. 7976, p. 1089–1100, 2023. Disponível em: <https://www.nature.com/articles/s41586-023-06415-8>. Acesso em: 30 de out. de 2023.

WEHNER, Christoph *et al.* Explainable Online Lane Change Predictions on a Digital Twin with a Layer Normalized LSTM and Layer-wise Relevance Propagation. *In: [S. l.: s. n.]*, 2022. p. 621–632. Disponível em: https://link.springer.com/10.1007/978-3-031-08530-7_52. Acesso em: 13 de nov. de 2023.

WEHNER, Christoph; KERTEL, Maximilian; WEWERKA, Judith. Interactive and Intelligent Root Cause Analysis in Manufacturing with Causal Bayesian Networks and Knowledge Graphs. *In: , 2023. 2023 IEEE 97th Vehicular Technology Conference (VTC2023-Spring)*. [S. l.]: IEEE, 2023. p. 1–7. Disponível em: <https://ieeexplore.ieee.org/document/10199563/>. Acesso em: 13 de nov. de 2023.

WHELAN, Ken E.; KING, Ross D. Intelligent software for laboratory automation. **Trends in Biotechnology**, [s. l.], v. 22, n. 9, p. 440–445, 2004. Disponível em: <https://linkinghub.elsevier.com/retrieve/pii/S0167779904002033>. Acesso em: 02 de nov. de 2023.

WICKSTROM, Lauren; GALLICCHIO, Emilio; LEVY, Ronald M. The linear interaction energy method for the prediction of protein stability changes upon mutation. **Proteins: Structure, Function, and Bioinformatics**, [s. l.], v. 80, n. 1, p. 111–125, 2012. Disponível em: <https://onlinelibrary.wiley.com/doi/10.1002/prot.23168>. Acesso em: 30 de out. de 2023.

WILBRAHAM, Liam; MEHR, S. Hessam M.; CRONIN, Leroy. Digitizing Chemistry Using the Chemical Processing Unit: From Synthesis to Discovery. **Accounts of Chemical Research**, [s. l.], v.

54, n. 2, p. 253–262, 2021. Disponível em: <https://pubs.acs.org/doi/10.1021/acs.accounts.0c00674>. Acesso em: 02 de nov. de 2023.

WILLIAMS, Kevin *et al.* Cheaper faster drug development validated by the repositioning of drugs against neglected tropical diseases. **Journal of The Royal Society Interface**, [s. l.], v. 12, n. 104, p. 20141289, 2015. Disponível em: <https://royalsocietypublishing.org/doi/10.1098/rsif.2014.1289>. Acesso em: 02 de nov. de 2023.

WONG, Felix *et al.* Benchmarking <sc>AlphaFold</sc> -enabled molecular docking predictions for antibiotic discovery. **Molecular Systems Biology**, [s. l.], v. 18, n. 9, 2022. Disponível em: <https://www.embopress.org/doi/10.15252/msb.202211081>. Acesso em: 02 de nov. de 2023.

WOODLEY, John M. Accelerating the implementation of biocatalysis in industry. **Applied Microbiology and Biotechnology**, [s. l.], v. 103, n. 12, p. 4733–4739, 2019. Disponível em: <http://link.springer.com/10.1007/s00253-019-09796-x>. Acesso em: 30 de out. de 2023.

WU, Gang *et al.* Artificial neural network approaches for fluorescence lifetime imaging techniques. **Optics Letters**, [s. l.], v. 41, n. 11, p. 2561, 2016. Disponível em: <https://opg.optica.org/abstract.cfm?URI=ol-41-11-2561>. Acesso em: 17 de abr. de 2023.

WU, Zachary *et al.* Machine learning-assisted directed protein evolution with combinatorial libraries. **Proceedings of the National Academy of Sciences**, [s. l.], v. 116, n. 18, p. 8852–8858, 2019. Disponível em: <https://pnas.org/doi/full/10.1073/pnas.1901979116>. Acesso em: 27 de mai. de 2023.

XIAO, Sian; VERKHIVKER, Gennady M.; TAO, Peng. Machine learning and protein allostery. **Trends in Biochemical Sciences**, [s. l.], 2022. Disponível em: <https://linkinghub.elsevier.com/retrieve/pii/S0968000422003115>. Acesso em: 24 de fev. de 2023.

XIE, Xuefeng *et al.* Comparison of random forest and multiple linear regression models for estimation of soil extracellular enzyme activities in agricultural reclaimed coastal saline land. **Ecological Indicators**, [s. l.], v. 120, p. 106925, 2021. Disponível em: <https://linkinghub.elsevier.com/retrieve/pii/S1470160X20308645>. Acesso em: 15 de nov. de 2023.

XU, Jia-qi *et al.* Cost-effectiveness Assessment of Commonly-used Drugs for Hepatoprotection and Enzymes Reduction Based on Decision Tree and Multi-utility Theory. **Chinese Herbal Medicines**, [s. l.], v. 6, n. 4, p. 305–312, 2014. Disponível em: <https://linkinghub.elsevier.com/retrieve/pii/S1674638414600460>. Acesso em: 15 de nov. de 2023.

XU, Chicheng *et al.* Machine learning in petrophysics: Advantages and limitations. **Artificial Intelligence in Geosciences**, [s. l.], v. 3, p. 157–161, 2022. Disponível em: <https://linkinghub.elsevier.com/retrieve/pii/S2666544122000338>. Acesso em: 13 de mar. de 2023.

XU, Xinhai *et al.* Supply chain finance: A systematic literature review and bibliometric analysis. **International Journal of Production Economics**, [s. l.], v. 204, p. 160–173, 2018. Disponível em: <https://linkinghub.elsevier.com/retrieve/pii/S0925527318303098>. Acesso em: 04 de mai. de 2023.

XU, Jia; VAN DOREN, Steven R. Affinities and Comparisons of Enzyme States by Principal Component Analysis of NMR Spectra, Automated using TREND Software. *In*: [S. l.: s. n.], 2018. p. 217–240. Disponível em: <https://linkinghub.elsevier.com/retrieve/pii/S0076687918302131>. Acesso em: 01 de out. de 2023.

YAHYAOU, A. *et al.* Contribution of artificial intelligence and big data in a medical biology laboratory: An experience of the central laboratory CHU Mohammed VI Oujda. **Materials Today: Proceedings**, [s. l.], v. 72, p. 3718–3723, 2023. Disponível em: <https://linkinghub.elsevier.com/retrieve/pii/S2214785322059442>. Acesso em: 13 de mar. de 2023.

YANG, Yahan *et al.* Expert recommendation on collection, storage, annotation, and management of data related to medical artificial intelligence. **Intelligent Medicine**, [s. l.], 2021. Disponível em: <https://linkinghub.elsevier.com/retrieve/pii/S2667102621001200>. Acesso em: 17 de abr. de 2023.

YANG, Shuxia *et al.* Modeling grassland above-ground biomass based on artificial neural network and remote sensing in the Three-River Headwaters Region. **Remote Sensing of Environment**, [s. l.], v. 204, p. 448–455, 2018. Disponível em: <https://linkinghub.elsevier.com/retrieve/pii/S0034425717304741>. Acesso em: 24 de fev. de 2023.

YING, Wang *et al.* Artificial Neural Network Prediction for Seismic Response of Bridge Structure. *In:* , 2009. **2009 International Conference on Artificial Intelligence and Computational Intelligence**. [S. l.]: IEEE, 2009. p. 503–506. Disponível em: <http://ieeexplore.ieee.org/document/5375889/>. Acesso em: 17 de abr. de 2023.

YUAN, Zengrui *et al.* Predicting mechanical behaviors of rubber materials with artificial neural networks. **International Journal of Mechanical Sciences**, [s. l.], v. 249, p. 108265, 2023. Disponível em: <https://linkinghub.elsevier.com/retrieve/pii/S0020740323001674>. Acesso em: 17 de abr. de 2023.

ZAREI, Esmaeil; KHAN, Faisal; ABBASSI, Rouzbeh. How to account artificial intelligence in human factor analysis of complex systems?. **Process Safety and Environmental Protection**, [s. l.], v. 171, p. 736–750, 2023. Disponível em: <https://linkinghub.elsevier.com/retrieve/pii/S0957582023000782>. Acesso em: 13 de mar. de 2023.

ZHANG, Yongchang *et al.* Big data and artificial intelligence based early risk warning system of fire hazard for smart cities. **Sustainable Energy Technologies and Assessments**, [s. l.], v. 45, p. 100986, 2021. Disponível em: <https://linkinghub.elsevier.com/retrieve/pii/S2213138820314144>. Acesso em: 13 de nov. de 2023.

ZHANG, Tao *et al.* Classification Models for Predicting Cytochrome P450 Enzyme-Substrate Selectivity. **Molecular Informatics**, [s. l.], v. 31, n. 1, p. 53–62, 2012. Disponível em: <https://onlinelibrary.wiley.com/doi/10.1002/minf.201100052>. Acesso em: 12 de mar. 2023.

ZHANG, Wentao *et al.* Heterogeneous catalysis mediated by light, electricity and enzyme via machine learning: Paradigms, applications and prospects. **Chemosphere**, [s. l.], v. 308, p. 136447, 2022. Disponível em: <https://linkinghub.elsevier.com/retrieve/pii/S004565352202940X>. Acesso em: 01 de out. de 2023.

ZHANG, Xiang-Long *et al.* Knowledge Domain and Emerging Trends in Vinegar Research: A Bibliometric Review of the Literature from WoSCC. **Foods**, [s. l.], v. 9, n. 2, p. 166, 2020. Disponível em: <https://www.mdpi.com/2304-8158/9/2/166>. Acesso em: 01 de out. de 2023.

ZHANG, R. *et al.* Modelling enzyme inhibition toxicity of ionic liquid from molecular structure via convolutional neural network model. **SAR and QSAR in Environmental Research**, [s. l.], p. 1–15, 2023. Disponível em: <https://www.tandfonline.com/doi/full/10.1080/1062936X.2023.2255517>. Acesso em: 14 de out. de 2023.

ZHANG, Fangrong *et al.* POEMS syndrome in the 21st century: A bibliometric analysis. **Heliyon**, [s. l.], v. 9, n. 10, p. e20612, 2023. Disponível em: <https://linkinghub.elsevier.com/retrieve/pii/S2405844023078209>. Acesso em: 13 de mar. de 2023.

ZHANG, Yun; KWONG, Sam; WANG, Shiqi. Machine learning based video coding optimizations: A survey. **Information Sciences**, [s. l.], v. 506, p. 395–423, 2020. Disponível em: <https://linkinghub.elsevier.com/retrieve/pii/S0020025519307145>. Acesso em: 13 de mar. de 2023.

ZHANG, Zhenjun; LIN, Jiajiang; CHEN, Zuliang. Predicting the effect of silver nanoparticles on soil enzyme activity using the machine learning method: type, size, dose and exposure time. **Journal of Hazardous Materials**, [s. l.], v. 457, p. 131789, 2023a. Disponível em: <https://linkinghub.elsevier.com/retrieve/pii/S0304389423010725>. Acesso em: 20 de ago. de 2023.

ZHANG, Lili; LING, Jie; LIN, Mingwei. Artificial intelligence in renewable energy: A comprehensive bibliometric analysis. **Energy Reports**, [s. l.], v. 8, p. 14072–14088, 2022. Disponível em: <https://linkinghub.elsevier.com/retrieve/pii/S2352484722022818>. Acesso em: 13 de nov. de 2023.

ZHAO, Yi *et al.* QSAR in natural non-peptidic food-related compounds: Current status and future perspective. **Trends in Food Science & Technology**, [s. l.], v. 140, p. 104165, 2023. Disponível em: <https://linkinghub.elsevier.com/retrieve/pii/S0924224423002807>. Acesso em: 18 de out. de 2023.

ZHENG, Jiahao *et al.* Exploiting machine learning for bestowing intelligence to microfluidics. **Biosensors and Bioelectronics**, [s. l.], v. 194, p. 113666, 2021. Disponível em: <https://linkinghub.elsevier.com/retrieve/pii/S095656632100703X>. Acesso em: 24 de fev. de 2023.

ZHONG, Ruizhi; SALEHI, Cyrus; JOHNSON, Ray. Machine learning for drilling applications: A review. **Journal of Natural Gas Science and Engineering**, [s. l.], v. 108, p. 104807, 2022. Disponível em: <https://linkinghub.elsevier.com/retrieve/pii/S1875510022003936>. Acesso em: 13 de mar. de 2023.

ZUBROD, Jochen P. *et al.* Physiological variables in machine learning QSARs allow for both cross-chemical and cross-species predictions. **Ecotoxicology and Environmental Safety**, [s. l.], v. 263, p. 115250, 2023. Disponível em: <https://linkinghub.elsevier.com/retrieve/pii/S0147651323007546>. Acesso em: 17 de nov. de 2023.