**ESTUDO BIBLIOMÉTRICO ENZIMAS MICROBIANAS NA BIOTECNOLOGIA INDUSTRIAL E SUAS TENDÊNCIAS.**

**BIBLIOMETRIC STUDY: MICROBIAL ENZYMES IN THE BIOCHEMICAL INDUSTRY AND THEIR TRENDS.**

David Lopes Maciel[[1]](#footnote-1)

Dr. Fabrício Moraes de Almeida[[2]](#footnote-2)

Natiele Vieira de Oliveira Maciel[[3]](#footnote-3)

Resumo –.

Palavras-chave:

*Abstract -.*

*Keywords:.*

**INTRODUÇÃO**

**FUNDAMENTAÇÃO TEÓRICA**

**METODOLOGIA**

Para a condução deste estudo bibliométrico sobre enzimas microbianas na indústria bioquímica, adotou-se o método PRIA-Processo de Revisão Integrado com Inteligência Artificial. Tal escolha fundamenta-se na necessidade de aliar rigor científico à eficiência operacional, assegurando a sistematização do levantamento, seleção e análise da produção científica. O PRIA organiza-se em cinco fases sequenciais, formulação da questão de pesquisa, busca sistemática, triagem automatizada, análise crítica e síntese, e redação final do manuscrito que, de forma integrada, permitem minimizar redundâncias, reduzir vieses e ampliar a robustez dos resultados obtidos.

Essa abordagem metodológica garante não apenas a rastreabilidade de cada etapa do processo investigativo, mas também a possibilidade de replicação em diferentes contextos científicos, consolidando-se como um modelo aplicável a estudos interdisciplinares que envolvem biotecnologia e bioquímica industrial.

A primeira etapa do método PRIA consiste na formulação da questão de pesquisa, elemento central que direciona todas as fases subsequentes do estudo. No contexto deste capítulo, buscou-se elaborar uma questão clara, objetiva e metodologicamente precisa, capaz de orientar o levantamento e a análise da produção científica internacional.

Assim, definiu-se como problema investigativo: “Quais são as tendências científicas e tecnológicas na aplicação de enzimas microbianas na indústria bioquímica no período de 2000 a 2024, considerando a evolução temporal, as redes de colaboração e os eixos temáticos predominantes?” Essa formulação estabelece não apenas o recorte temporal e temático da investigação, mas também delimita os resultados esperados, os quais incluem a identificação de núcleos de conhecimento, de periódicos de maior impacto e das lacunas existentes na literatura.

Após a formulação inicial, a questão de pesquisa foi submetida a um processo de refinamento metodológico, conforme previsto no método PRIA. Essa etapa teve como objetivo garantir maior clareza conceitual, precisão nos limites temporais e temáticos, além de alinhamento com os critérios bibliométricos. O processo envolveu a revisão iterativa da questão formulada, apoiada por ferramentas digitais de inteligência artificial, utilizadas para avaliar a consistência semântica dos termos e a sua aderência ao escopo da investigação. Esse procedimento permitiu reduzir ambiguidades, eliminar redundâncias e assegurar que os descritores principais fossem compatíveis com as bases de dados selecionadas.

Com base nesse refinamento, a questão final manteve a essência investigativa, mas foi ajustada em sua forma, de modo a ampliar a objetividade analítica: “Quais são as tendências científicas e tecnológicas na aplicação de enzimas microbianas na indústria bioquímica entre os anos de 2000 e 2024, considerando indicadores de evolução temporal, redes de colaboração científica e núcleos temáticos emergentes?” Essa versão estruturada, mais próxima da lógica bibliométrica, atende à recomendação de Garfield (2006) de que a formulação da pergunta deve orientar todo o processo de análise e sistematização da produção científica. Além disso, alinha-se à orientação de Aria e Cuccurullo (2017), segundo os quais a qualidade das análises bibliométricas depende diretamente da clareza e da operacionalidade da questão de pesquisa.

A operacionalização prática consistiu em um processo iterativo em três momentos. Primeiro, a questão formulada foi testada em consultas exploratórias nas bases Web of Science e Scopus, de modo a identificar a amplitude e a relevância dos resultados iniciais. Em segundo lugar, os termos recuperados foram comparados com vocabulários controlados e indexadores das próprias bases, permitindo a inclusão de sinônimos e a exclusão de termos pouco representativos. Por fim, a versão refinada da questão foi consolidada a partir de critérios de precisão (recuperar estudos efetivamente relacionados ao tema) e de abrangência (evitar perda de trabalhos relevantes). Essa operacionalização prática assegurou que a questão final fosse não apenas conceitualmente robusta, mas também tecnicamente aplicável à estratégia de busca sistemática que orientará as fases subsequentes do PRIA.

Com a questão de pesquisa refinada, procedeu-se à definição da estratégia de busca estruturada, etapa fundamental para assegurar a rastreabilidade e a reprodutibilidade do estudo. Foram selecionadas as bases de dados Web of Science (WoS), Scopus e PubMed, reconhecidas internacionalmente pela cobertura ampla e multidisciplinar em biotecnologia, bioquímica e ciências aplicadas. O recorte temporal estabelecido compreendeu o período de 2000 a 2024, em conformidade com a delimitação da questão investigativa, garantindo tanto a análise de tendências históricas quanto a identificação de avanços recentes.

A formulação da string de busca baseou-se em operadores booleanos e combinações de termos livres e controlados. Foram utilizados descritores principais como “microbial enzymes”, “industrial enzymes” e “enzymatic biocatalysis”, associados a termos de aplicação, como “biochemical industry”, “bioprocessing” e “biotechnology applications”. Assim, a estratégia típica assumiu a forma: (“microbial enzymes” OR “industrial enzymes” OR “enzymatic biocatalysis”) AND (“biochemical industry” OR “bioprocessing” OR “biotechnology applications”). Esse conjunto de descritores foi validado por meio de buscas-piloto, assegurando que recuperasse um corpus representativo e relevante, sem comprometer a precisão.

Os critérios de inclusão compreenderam artigos originais revisados por pares, publicados em inglês ou português, com foco explícito na aplicação de enzimas microbianas em processos industriais bioquímicos. Excluíram-se revisões narrativas sem base empírica, documentos de opinião, resumos de conferências e publicações sem acesso integral. A execução da busca estruturada resultou em um conjunto inicial de registros que será submetido à etapa subsequente de triagem automatizada, conforme preconizado pelo método PRIA.

Na etapa inicial da busca estruturada, a aplicação da string preliminar resultou na recuperação de apenas 27 documentos, número insuficiente para compor um corpus bibliométrico robusto. Esse resultado evidenciou a necessidade de ajustes nos descritores e na amplitude da consulta. A partir desse diagnóstico, procedeu-se ao refinamento da estratégia, com a inclusão de sinônimos em inglês e português, além da expansão dos termos relacionados às aplicações industriais. Como consequência, a aplicação da nova string na base Web of Science Core Collection elevou o número de registros para 101 documentos, garantindo maior representatividade temática e temporal. Essa diferença demonstra a importância da etapa iterativa do PRIA, na qual sucessivas rodadas de testes e ajustes são essenciais para alcançar equilíbrio entre abrangência e precisão. Link da Pesquisa: https://www.webofscience.com/wos/woscc/summary/6da7a5c4-3ec3-48fc-b077-92cb64f8203b-0174beb54c/relevance/1

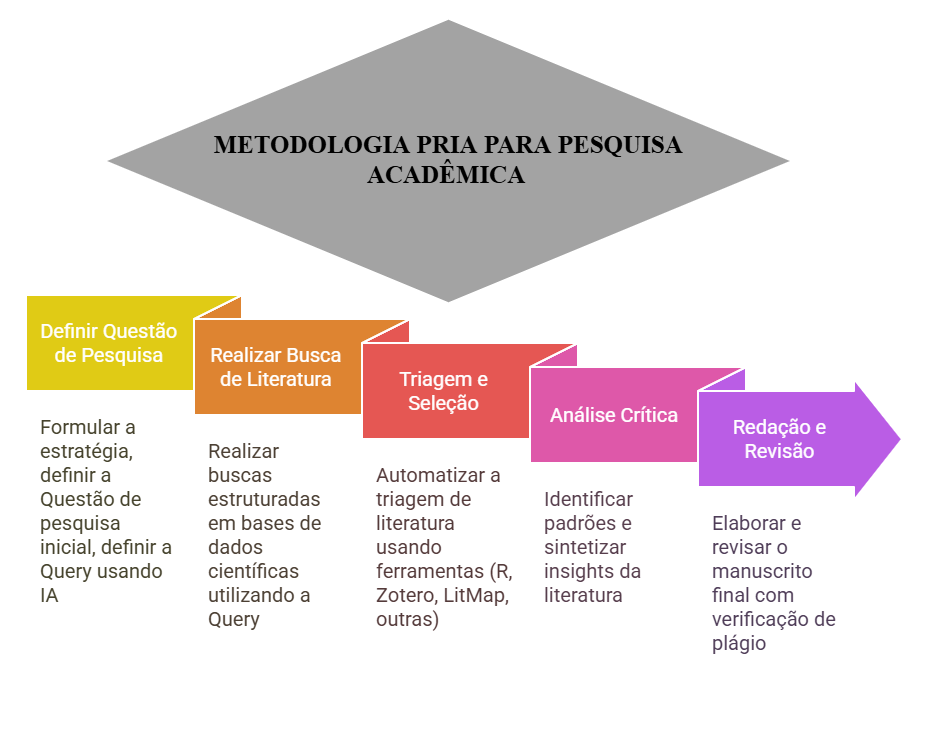
Complementarmente à busca realizada na Web of Science, procedeu-se à aplicação da estratégia refinada também na base Scopus, a fim de ampliar a representatividade do corpus bibliográfico. Para compatibilizar a sintaxe entre as plataformas, a string foi adaptada para o formato próprio da Scopus, utilizando o campo TITLE-ABS-KEY, de modo a recuperar registros nos títulos, resumos e palavras-chave. A versão final da consulta foi estruturada da seguinte forma: TITLE-ABS-KEY (“enzimas microbianas” OR “enzimas industriais” OR “biocatálise enzimática” OR “microbial enzymes” OR “industrial enzymes” OR “enzymatic biocatalysis” OR “enzyme technology” OR “biorefinery enzymes”) AND TITLE-ABS-KEY (“indústria bioquímica” OR “bioprocessamento” OR “biotecnologia industrial” OR “biochemical industry” OR “bioprocessing” OR “industrial biotechnology” OR “biorefinery”). Foram aplicados filtros para o período de 2000 a 2024, considerando artigos originais e de revisão publicados em inglês e português. Esse procedimento assegurou consistência metodológica entre as bases e contribuiu para ampliar a abrangência dos registros analisados.

Na base Scopus, a aplicação da string refinada e adaptada à sua sintaxe resultou na recuperação de 131 documentos, abrangendo o período de 2000 a 2024. Foram incluídos artigos originais e de revisão, publicados em inglês e português, mantendo-se a mesma lógica de filtros empregada na Web of Science. Esse número expressivo confirma a relevância do tema no contexto industrial e bioquímico, ampliando a representatividade do corpus bibliográfico. A integração dos resultados obtidos na Scopus com os registros da Web of Science assegura maior cobertura de periódicos e de áreas interdisciplinares, consolidando uma base mais robusta para as etapas de triagem e análise bibliométrica subsequentes.

Complementando as buscas realizadas na Web of Science e Scopus, a estratégia foi igualmente adaptada e aplicada à base PubMed, a fim de contemplar a produção científica indexada nas áreas de biotecnologia, microbiologia e ciências da saúde. Para compatibilizar a sintaxe da consulta, utilizou-se o campo [tiab], de forma a recuperar os termos de interesse em títulos e resumos. A string incluiu sinônimos em inglês e português, assegurando consistência com as demais bases. Foram aplicados filtros para o período de 2000 a 2024, abrangendo artigos originais e de revisão em inglês e português. Essa etapa reforçou a abrangência interdisciplinar do corpus, garantindo que estudos relevantes publicados em periódicos biomédicos também fossem incorporados ao universo bibliográfico a ser submetido à triagem.

Em seguida apresentamos o fluxo de atividades do método PRIA, conforme a figura XX

Figura XX – Metodologia PRIA



Fonte: autor, 2025

Na terceira fase do método PRIA, procedeu-se à integração dos registros recuperados nas três bases de dados em um corpus único de análise. Para tanto, os metadados foram exportados em seus formatos originais: BibTeX (.bib) para Web of Science e Scopus, e NBIB (.nbib) para PubMed. Os arquivos foram importados no pacote Bibliometrix, no ambiente RStudio, por meio da função convert2df, que permite a conversão padronizada dos registros em formato tabular. Esse procedimento assegurou a unificação de 101 registros da Web of Science, 131 da Scopus e 66 do PubMed, totalizando 298 entradas brutas. Em seguida, iniciou-se a etapa de triagem, que compreendeu a detecção e remoção de duplicatas, bem como a aplicação de critérios de inclusão e exclusão previamente definidos. Esse processo garantiu a consistência do corpus bibliográfico, estabelecendo a base definitiva para as análises bibliométricas subsequentes.

O processamento dos dados bibliográficos iniciou-se com a configuração do ambiente analítico no RStudio, contemplando a instalação do pacote Bibliometrix e de suas dependências. Após a preparação do ambiente, os arquivos exportados das bases de dados foram importados utilizando a função convert2df, que converte os registros em um formato tabular unificado. Esse procedimento permitiu a integração dos metadados provenientes da Web of Science, Scopus e PubMed, resultando em uma base consolidada, livre de duplicatas e devidamente padronizada para análises subsequentes.

Em seguida, os registros processados foram exportados em formato Excel (.xlsx), facilitando o controle manual e a conferência dos dados. A versão final da base, já revisada, foi posteriormente importada no Biblioshiny, interface gráfica do Bibliometrix, onde foram conduzidas as análises bibliométricas interativas. Esse fluxo metodológico — instalação do ambiente, geração da base unificada, exportação e tratamento no Excel, seguido do carregamento no Biblioshiny — assegurou rastreabilidade, transparência e reprodutibilidade ao estudo, consolidando as bases empíricas necessárias para a exploração das tendências científicas e tecnológicas relacionadas às enzimas microbianas na indústria bioquímica.

**RESULTADOS**

Durante a etapa de mapeamento estrutural das bases de dados, foi realizada uma análise comparativa das variáveis contidas nos arquivos oriundos das plataformas PubMed, Scopus e Web of Science (WoS). O objetivo central foi identificar as colunas comuns entre os conjuntos de dados, garantindo compatibilidade e integridade na construção da base unificada a ser utilizada nas análises bibliométricas.

Como resultado, constatou-se que apenas 25 variáveis estavam simultaneamente presentes nas três bases, compondo um núcleo padronizado mínimo. Dentre essas, destacam-se: AU (autores), TI (título), AB (resumo), PY (ano de publicação), SO (fonte), DI (DOI), CR (referências citadas), DE (descritores dos autores), TC (total de citações), ID (descritores indexados), DT (tipo de documento), J9 (abreviação da fonte), RP (autor correspondente), VL (volume), PP (página inicial), AF (nome completo dos autores), C1 (afiliações institucionais), LA (idioma), SR (referência abreviada), SR\_FULL (referência completa), AB\_raw, DE\_raw, KW\_Merged, DB (base de origem) e TI\_raw.

Por outro lado, verificou-se a existência de variáveis exclusivas de cada base, muitas delas relacionadas a identificadores internos, formatos de indexação ou campos redundantes. No caso da PubMed, foram identificadas como exclusivas as variáveis PMID, PMC, MHDA, PHST, STAT, EDAT, VI, PG, FAU e JT. Já na Scopus, destacaram-se como exclusivas as colunas Page.start, Page.end, URL, OA, Correspondence.Address, Correspondence.E-mail, Author.Keywords, Funding.Details e Article.Type. Por fim, a base Web of Science apresentou variáveis únicas como U2, usage.count.last.180.days, usage.count.since.2013, Research.Areas, Categories, Z9, C2, PU, BP, EI, FX, GP, NR, SC e SN.

Diante desse cenário, optou-se por considerar, na etapa de unificação, apenas o conjunto de variáveis comuns às três bases, visando assegurar a interoperabilidade e robustez das análises no ambiente do pacote Bibliometrix e sua interface gráfica Biblioshiny. As variáveis exclusivas foram preservadas separadamente para eventuais análises complementares ou consulta documental futura, mas não foram incorporadas à base consolidada principal.

Durante a inspeção preliminar das bases importadas, procedeu-se à identificação de registros bibliográficos desprovidos de identificador digital (DOI), elemento essencial para rastreamento, recuperação e reclassificação posterior dos artigos. A partir da filtragem automatizada, constatou-se a presença de 19 artigos sem DOI, distribuídos entre as três bases analisadas: PubMed (8), Web of Science (4) e Scopus (7). Os registros foram consolidados em uma planilha específica, contendo os campos essenciais para futura recuperação manual ou semiautomática, incluindo título, autores, ano de publicação, periódico e ID interno. Os arquivos, após leitura foram excluídos da base por não estarem dentro do contexto do tema.

Durante a etapa de refinamento das bases bibliográficas, foi aplicado um protocolo de pré-processamento sistemático, fundamental para garantir a integridade, a padronização e a compatibilidade dos dados antes da unificação. Inicialmente, identificaram-se os registros sem DOI, os quais foram marcados com uma variável de controle (NO\_DOI) para fins de rastreabilidade, embora não tenham sido removidos automaticamente, respeitando-se seu potencial informativo em análises temáticas. Em seguida, eliminou-se duplicações internas com base na combinação dos campos de título e identificador digital, preservando apenas a primeira ocorrência.

TABELA DE CRITÉRIOS DE PRÉ-PROCESSAMENTO BIBLIOMÉTRICO

| Nº | Critério | Descrição técnica | Justificativa metodológica |
| --- | --- | --- | --- |
| 1 | Identificação de registros sem DOI | Criação da coluna NO\_DOI, marcada como TRUE quando DI está ausente ou vazio | Permite rastreabilidade e decisão analítica posterior, sem exclusão prematura |
| 2 | Criação da coluna DB (base de origem) | Adiciona campo "PUBMED", "WOS" ou "SCOPUS" a todos os registros | Garante controle, segmentação e rastreamento pós-unificação |
| 3 | Remoção de duplicatas internas | Exclusão de registros com mesma combinação TI + DI | Elimina redundância na análise de produção e citação |
| 4 | Criação de campos essenciais ausentes | Adiciona campos obrigatórios inexistentes (AU, TI, SO, DI, PY, CR, etc.) com valor "ND" | Garante compatibilidade com funções do bibliometrix e evita falhas de execução |
| 5 | Preenchimento seletivo de valores nulos | Substituição de NA e "" por "ND" apenas quando o campo está vazio | Previne erros em cálculos, mantendo dados originais intactos |
| 6 | Sanitização textual de campos bibliométricos | Remoção de quebras, tabs, espaços extras, e padronização para codificação UTF-8 | Evita ruídos em análises textuais, coautoria e co-palavras |
| 7 | Validação do ano de publicação (PY) | Conversão para character e filtragem por padrão YYYY, senão substituição por "ND" | Evita distorções em curvas temporais e segmentações por período |
| 8 | Normalização de autores (AU) e fontes (SO) | Conversão para caixa alta (toupper) preservando acentuação | Garante consistência em redes de coautoria e análise por periódicos |
| 9 | Preservação dos dados originais válidos | Nenhum campo com valor válido é sobrescrito; alterações ocorrem apenas quando ausente ou inconsistente | Mantém fidelidade da base original e transparência metodológica |
| 10 | Inclusão de campos adicionais compatíveis com o Biblioshiny | Adição de campos como DT, AB, C1, RP, ID, TC, entre outros | Garante pleno funcionamento da interface biblioshiny e uso de todas as funcionalidades |
| 11 | Reorganização da estrutura das colunas | Ordenação padronizada das colunas principais à frente, conforme documentação de uso | Facilita visualização, exportação e compatibilidade entre bases |

A análise bibliométrica foi operacionalizada por meio do uso do pacote Bibliometrix, em sua interface gráfica Biblioshiny, com o objetivo de explorar de maneira interativa as estruturas conceitual, intelectual e social do corpus bibliográfico unificado. Para garantir a integridade do processo, os dados foram previamente tratados e consolidados em um único arquivo no formato .xlsx, denominado base\_bibliometrica\_unificada.xlsx, armazenado localmente no diretório do projeto. O Biblioshiny não realiza conexões com bases de dados externas (tais como PubMed, Scopus ou Web of Science); dessa forma, todas as análises dependem exclusivamente do conjunto de dados fornecido pelo usuário.

Durante o processo de inicialização, a base consolidada foi manualmente carregada por meio da opção “Import File”, disponível no painel principal da interface. A partir disso, as funcionalidades da aplicação passaram a operar sobre esse arquivo local, o qual já continha os campos bibliográficos essenciais para execução dos cálculos e geração de visualizações. Tal abordagem assegura rastreabilidade, reprodutibilidade e controle metodológico sobre as informações utilizadas nas análises subsequentes.

Para a etapa de análise interativa dos dados, utilizou-se a interface gráfica Biblioshiny, pertencente ao pacote bibliometrix em ambiente R. A aplicação foi executada localmente após a configuração do ambiente virtual por meio do gerenciador renv, garantindo reprodutibilidade e isolamento dos pacotes utilizados. Para iniciar a interface, foi necessário carregar explicitamente a biblioteca com library(bibliometrix), seguida do comando biblioshiny(). Esta ação iniciou um servidor local acessível via navegador, no qual foram realizadas todas as análises exploratórias sobre a base bibliográfica unificada.

Durante o processo de ativação, eventuais dependências complementares foram automaticamente instaladas, assegurando o pleno funcionamento da aplicação. Ressalta-se que o Biblioshiny opera exclusivamente com arquivos locais, previamente estruturados e validados, não realizando conexões externas com bases de dados científicas. Tal configuração assegura total controle sobre o corpus analisado e reforça os critérios de transparência e rastreabilidade científica exigidos pela metodologia PRIA adotada neste estudo.

Após a unificação das bases de dados provenientes da PubMed, Web of Science e Scopus, com um total consolidado de 298 artigos, foi realizada uma etapa crítica de verificação quanto à compatibilidade do arquivo resultante para uso na interface Biblioshiny, do pacote Bibliometrix. Durante o processo de exportação para o formato .xlsx, foi detectado um erro relacionado ao limite de caracteres por célula imposto pelo Excel (32.767 caracteres). A análise identificou que apenas uma célula, localizada na linha 231 da coluna CR (referências citadas), ultrapassava esse limite, com um total de 36.364 caracteres.

Para contornar esse obstáculo sem comprometer a integridade geral da base, o conteúdo dessa célula foi truncado, limitando-se aos 200 primeiros caracteres. Essa medida foi suficiente para possibilitar a exportação correta do arquivo .xlsx, sem violar os padrões técnicos da biblioteca writexl.

Posteriormente, foi realizada uma análise de consistência da estrutura da base unificada, verificando:

a existência de colunas totalmente vazias;

a presença de campos obrigatórios;

a padronização de nomes de variáveis compatíveis com os requisitos do Bibliometrix/Biblioshiny;

a estrutura UTF-8 do arquivo e integridade geral dos dados.

Verificou-se que, com exceção da célula previamente ajustada, todas as demais colunas estavam adequadamente estruturadas, e não havia inconsistências impeditivas. O arquivo foi salvo no diretório oficial do projeto com o nome:

base\_bibliometrica\_unificada.xlsx

Foi também confirmado que o nome do arquivo não interfere na importação para o Biblioshiny, desde que:

o formato esteja correto (.xlsx, .bib, ou .csv),

a primeira linha contenha os nomes das colunas,

e os campos estejam devidamente padronizados.

Com isso, concluiu-se que a base está pronta para ser carregada no Biblioshiny para a execução das análises bibliométricas.

**CONCLUSÃO**

**REFERENCIAS**

**COPYRIGHT**

Direitos autorais: Os autores são os únicos responsáveis pelo material incluso no artigo.

1. Master in Master of Science in Emergent Technologies in Education. MUST UNIVERSITY, MUST, EUA, Academic of the Doctoral Program in Regional Development and Environment (PGDRA/UNIR). E-mail: [maciel.1000@hotmail.com](mailto:maciel.1000@hotmail.com).

   2 PhD in Physics (UFC), with post-doctorate in Scientific Regional Development (DCR/CNPq). Researcher of the Doctoral and Master Program in Regional Development and Environment (PGDRA/UNIR). E-mail: [dr.fabriciomoraes001@gmail.com](mailto:dr.fabriciomoraes001@gmail.com).

   3 Especialista em Oncologia - Centro Universitário - UNINTER, Especialista em Saúde da Família – Instituto Federal do Amazonas, Graduação em Enfermagem, União das Escolas Superiores de Rondônia – UNIRON. E-mail: [Natiele.vieira.oliveira@gmail.com](mailto:Natiele.vieira.oliveira@gmail.com) [↑](#footnote-ref-1)
2. [↑](#footnote-ref-2)
3. [↑](#footnote-ref-3)