



UNIVERSIDADE FEDERAL DE PERNAMBUCO
CENTRO DE INFORMÁTICA
SISTEMAS DE INFORMAÇÃO

GABRIEL CAVALCANTI DE MELO

**IDENTIFICAÇÃO DE EXOPLANETAS UTILIZANDO INTELIGÊNCIA
ARTIFICIAL: UMA REVISÃO SISTEMÁTICA DE LITERATURA**

Recife

2022

GABRIEL CAVALCANTI DE MELO

**IDENTIFICAÇÃO DE EXOPLANETAS UTILIZANDO INTELIGÊNCIA
ARTIFICIAL: UMA REVISÃO SISTEMÁTICA DE LITERATURA**

Trabalho apresentado ao Programa de Graduação em Sistemas de Informação do Centro de Informática da Universidade Federal de Pernambuco como requisito parcial para obtenção do grau de Bacharel em Sistemas de Informação.

Orientador: Prof. Dr. Fernando Maciano de Paula Neto

Recife

2022

GABRIEL CAVALCANTI DE MELO

**IDENTIFICAÇÃO DE EXOPLANETAS UTILIZANDO INTELIGÊNCIA
ARTIFICIAL: UMA REVISÃO SISTEMÁTICA DE LITERATURA**

Trabalho apresentado ao Programa de Graduação em Sistemas de Informação do Centro de Informática da Universidade Federal de Pernambuco como requisito parcial para obtenção do grau de Bacharel em Sistemas de Informação.

Recife, 12 de Maio de 2022

BANCA EXAMINADORA

Prof. Fernando Maciano de Paula Neto (Orientador)

UNIVERSIDADE FEDERAL DE PERNAMBUCO

Prof. Jessyka Flavyanne Ferreira Vilela (2º membro da banca)

UNIVERSIDADE FEDERAL DE PERNAMBUCO

AGRADECIMENTOS

Gostaria de agradecer a todos que me ajudaram neste processo. Aos professores, em especial ao professor Fernando Neto que fez parte do primeiro contato que eu tive com programação, no início do curso, e agora faz parte da conclusão, sendo o orientador deste trabalho. Obrigado pela ajuda e incentivo em todo esse processo.

A minha família, minha irmã e em especial aos meus pais que sempre me apoiaram para que eu pudesse conseguir chegar até aqui. Obrigado pelo incentivo, ensinamentos, apoio, carinho e todo o esforço que sempre fizeram. Vocês são importantes demais e dedico tudo isso a vocês. Obrigado por terem sempre colocado a minha educação em primeiro lugar.

À minha companheira de vida, Talyta, que tanto me ajudou neste processo, me dando força, apoio, incentivo e enfrentando ao meu lado os desafios e o dia a dia da graduação. Obrigado por me ensinar tantas coisas, por me incentivar e não me deixar ter medo de tentar. Te conhecer neste curso e termos passado por tudo isso juntos me deixa muito feliz e grato. Com certeza, ter você ao meu lado tornou tudo mais fácil. Obrigado demais por tudo!

*“Em algum lugar, algo incrível está
esperando para ser descoberto.”*

(Carl Sagan)

RESUMO

Até o ano de 1992, não existiam provas da existência de um planeta orbitando uma outra estrela que não fosse o Sol. Os métodos manuais utilizados inicialmente para a descoberta dos primeiros exoplanetas eram trabalhosos e estavam sujeitos a erros. Porém, com o auxílio da Inteligência artificial (IA) para analisar os dados astronômicos provenientes das missões espaciais, o número de exoplanetas cresce a cada ano, chegando até a quantidade de milhares de exoplanetas confirmados. Diante deste cenário, o trabalho visa fazer uma revisão sistemática de literatura (RSL) do uso da Inteligência Artificial na identificação de exoplanetas, para que tenhamos, em um só trabalho, uma síntese das informações sobre quais técnicas inteligentes estão sendo utilizadas, quais são suas contribuições e impactos, e quais as limitações e obstáculos enfrentados pelo uso de métodos inteligentes na área de identificação de exoplanetas.

Palavras-chave: Inteligência artificial; detecção de exoplanetas; identificação de exoplanetas; aprendizado de máquina; aprendizado profundo.

ABSTRACT

Until 1992, there was no evidence of the existence of a planet orbiting a star other than the Sun. The manual methods used initially for detecting the firsts exoplanets were laborious and with a high chance of making mistakes. However, with the help of Artificial Intelligence (AI) to analyze astronomical data from space missions, the number of exoplanets grows every year, reaching the number of thousands of confirmed exoplanets. Given this scenario, this work aims to make a systematic literature review (RSL) on the use of Artificial Intelligence in the identification of exoplanets, so that we have, in a single work, a synthesis of information about which intelligent techniques are being used, which are their contributions and impacts, and what are the limitations and obstacles faced by the use of intelligent methods in the area of exoplanet identification.

Keywords: Artificial intelligence; exoplanet detection; exoplanet identification; machine learning; deep learning.

LISTA DE FIGURAS

Figura 1 — Curva de Luz do primeiro exoplaneta descoberto pelo método de trânsito	16
Figura 2 — Exemplo de Imagem Direta de um exoplaneta	17
Figura 3 — Detalhamento das etapas do método proposto por Kitchenham	19
Figura 4 — Estudos identificados em cada base de busca	26
Figura 5 — Estudos selecionados após aplicação de cada critério	27
Figura 6 — Estudos selecionados em cada base	28
Figura 7 — Utilização dos métodos de detecção	41

LISTA DE TABELAS

Tabela 1 — Bases de dados utilizadas	21
Tabela 2 — Palavra-chave e seus respectivos sinônimos	22
Tabela 3 — String de busca da pesquisa	22
Tabela 4 — Critérios de exclusão	23
Tabela 5 — Critérios de inclusão	23
Tabela 6 — Classificação dos estudos de acordo com a média obtida	25
Tabela 7 — ID dos estudos selecionados	28
Tabela 8 — Resultados da avaliação de qualidade	31
Tabela 9 — Técnicas utilizadas nos estudos	35
Tabela 10 — Métodos utilizados para detecção	38
Tabela 11 — Técnicas utilizadas para cada método de detecção	41
Tabela 12 — Técnicas com melhor desempenho	42
Tabela 13 — Impactos e contribuições	47
Tabela 14 — Técnicas utilizadas em cada contribuição e impacto identificado	51
Tabela 15 — Limites e Desafios	53
Tabela 16 — Técnicas utilizadas em cada desafio ou limitação identificado	57

LISTA DE ABREVIATURAS E SIGLAS

IA Inteligência artificial

RSL Revisão Sistemática de Literatura

CE Critérios de exclusão

CI Critérios de inclusão

CNN *Convolutional Neural Network*

ES Estudos selecionados

SUMÁRIO

1	INTRODUÇÃO	12
1.1	OBJETIVOS	13
1.1.1	OBJETIVOS GERAIS	13
1.1.2	OBJETIVOS ESPECÍFICOS	13
1.2	PERGUNTA DE PESQUISA	13
1.3	TRABALHOS RELACIONADOS	14
2	REFERENCIAL TEÓRICO	15
2.1	EXOPLANETAS	15
2.2	MÉTODOS DE DETECÇÃO	15
2.2.1	VELOCIDADE RADIAL	15
2.2.2	TRANSITO	16
2.2.3	IMAGEM DIRETA	17
2.3	INTELIGÊNCIA ARTIFICIAL	18
3	METODOLOGIA	18
3.1	REVISÃO SISTEMÁTICA DE LITERATURA	18
3.2	ETAPAS DA REVISÃO	18
4	PROTOCOLO DA REVISÃO SISTEMÁTICA DA LITERATURA	20
4.1	PERGUNTAS DA PESQUISA	20
4.2	ESTRATÉGIA DE BUSCA	21
4.2.1	FONTES DE BUSCA	21
4.2.2	TERMOS DE BUSCA	22
4.2.3	STRING DE BUSCA	22
4.3	CRITÉRIOS DE SELEÇÃO DOS ESTUDOS	23
4.3.1	CRITÉRIOS DE EXCLUSÃO	23
4.3.2	CRITÉRIOS DE INCLUSÃO	23
4.4	PROCEDIMENTOS DA SELEÇÃO DOS ESTUDOS	24
4.5	CRITÉRIOS DA AVALIAÇÃO DE QUALIDADE	24
4.6	ESTRATÉGIA DA EXTRAÇÃO DOS DADOS	25
5	SELEÇÃO DOS ESTUDOS E APRESENTAÇÃO DOS RESULTADOS	25
5.1	IDENTIFICAÇÃO DOS ESTUDOS	25
5.2	SELEÇÃO DOS ESTUDOS	26
5.3	AVALIAÇÃO DE QUALIDADE	28
5.4	EXTRAÇÃO DOS DADOS	33
5.5	APRESENTAÇÃO E SÍNTESE DOS RESULTADOS	34
5.5.1	Q1: QUAIS TÉCNICAS INTELIGENTES SÃO UTILIZADAS ATUALMENTE PARA IDENTIFICAR EXOPLANETAS?	34
5.5.2	Q2: QUAIS MÉTODOS DE IDENTIFICAÇÃO DE EXOPLANETAS TIVERAM MAIOR CONTRIBUIÇÃO DO USO DE INTELIGÊNCIA ARTIFICIAL?	38
5.5.3	Q3: QUAIS TÉCNICAS APRESENTARAM MELHOR DESEMPENHO?	43

5.5.4	Q4: QUAIS OS PRINCIPAIS IMPACTOS E CONTRIBUIÇÕES QUE ESSAS TÉCNICAS TROUXERAM PARA A IDENTIFICAÇÃO DE EXOPLANETAS?	47
5.5.5	Q5: QUAIS OS PRINCIPAIS DESAFIOS OU LIMITAÇÕES QUE AS TÉCNICAS APLICADAS NOS ESTUDOS ENFRENTARAM EM SEUS CONTEXTOS?	53
5.5.6	LIMITAÇÕES DA PESQUISA E AMEAÇA À VALIDADE	58
6	CONCLUSÃO E TRABALHOS FUTUROS	58
6.1	CONCLUSÃO	58
6.2	TRABALHOS FUTUROS	60
	REFERÊNCIAS	61

1 INTRODUÇÃO

Antes da descoberta do primeiro exoplaneta, a existência de planetas orbitando outras estrelas que não fossem o sol era apenas uma hipótese, que tinha grande aceitação, pois fazia sentido que existissem, já que o nosso sol não poderia ser o único a possuir planetas, dentre uma infinidade de outras estrelas no universo. Porém, detectar e provar a existência desses objetos era uma tarefa difícil, já que exoplanetas não emitem luz própria e estão muito distantes [1]. Porém, tudo mudou quando, em 1992, Aleksander Wolszczan provou, pela primeira vez, a existência de um planeta orbitando a estrela PSR B1257+12 [2].

Já se passaram 30 anos desde a primeira confirmação, e até o momento foram descobertos mais de 5000 exoplanetas [3]. Porém, os métodos manuais utilizados inicialmente para analisar candidatos a exoplanetas eram trabalhosos e facilmente suscetíveis a erros humanos [4]. Além disso, a quantidade de dados de observações espaciais cresceu rapidamente com o passar dos anos, principalmente devido aos lançamentos de missões espaciais dedicadas à busca por exoplanetas, que se iniciou em 2006, quando a missão *CoRoT* foi lançada pela agência espacial francesa.

Já o lançamento da missão *Kepler* em 2009 fez com que o número de exoplanetas encontrados crescesse rapidamente [6]. O mesmo aconteceu com o lançamento da *Transiting Exoplanet Survey Satellite* (TESS), a mais nova missão espacial focada na busca por exoplanetas da NASA que foi lançada em 2018, e é capaz de analisar dados de estrelas que as outras missões não eram capazes [7]. Porém, não só o número de exoplanetas confirmados aumentou, mas também o número de candidatos, que já somam mais de 8000 de acordo com o catálogo de exoplanetas da NASA [8]. Tal catálogo reúne dados de estrelas das mais variadas composições, tamanhos e magnitudes de brilho.

Neste trabalho, mostraremos, através de uma RSL, que a inteligência artificial é capaz de analisar esta grande quantidade de dados, com o objetivo de detectar e confirmar a existência de exoplanetas, de forma mais rápida e menos sujeita a erros. Por fim, mostraremos as principais contribuições do uso de IA na identificação de exoplanetas e até onde seu uso pode chegar nessa atividade, ao mostrar quais os limites e desafios presentes nesse contexto.

1.1 OBJETIVOS

1.1.1 OBJETIVOS GERAIS

O objetivo deste trabalho é realizar uma revisão sistemática de literatura do uso da inteligência artificial na área da astronomia, mais especificamente no ramo da identificação de exoplanetas. De modo que possamos compreender o cenário atual, e conseqüentemente, saibamos como a inteligência artificial contribui para essa área da ciência e como ainda pode contribuir.

1.1.2 OBJETIVOS ESPECÍFICOS

- Realizar uma busca nas principais bases acadêmicas estudos que abordam o uso de Inteligência Artificial que auxiliam na identificação de exoplanetas;
- Identificar quais ferramentas e ramos da inteligência artificial estão sendo utilizados para a identificação de exoplanetas;
- Identificar quais os avanços e melhorias o uso da inteligência artificial trouxe para a identificação de exoplanetas;
- Identificar quais os limites e desafios do uso da IA para a área de identificação de exoplanetas.

1.2 PERGUNTA DA PESQUISA

A pergunta geral que este trabalho visa responder é: **“Quais as características, contribuições e limites das técnicas de inteligência artificial que contribuem para a identificação de exoplanetas?”**

1.3 TRABALHOS RELACIONADOS

Não foram encontrados trabalhos de revisão de literatura acerca do uso de Inteligência artificial para identificação de exoplanetas, mas foram encontrados trabalhos de revisão de literatura que abordam temas envolvendo inteligência artificial, como por exemplo o trabalho de Nascimento et al (2020), que faz uma revisão sistemática de literatura (RSL) sobre a engenharia de software para inteligência artificial e aprendizado de máquina. Watson et al (2020) faz uma RSL sobre o uso de aprendizado profundo na pesquisa de engenharia de software.

Portanto vimos que existem trabalhos de revisão sistemática de literatura na área de inteligência artificial, porém nenhum deles se especificou nas técnicas de IA utilizadas para identificação de exoplanetas. Diante deste cenário, este trabalho visa suprir essa lacuna.

2 REFERENCIAL TEÓRICO

2.1 EXOPLANETAS

Exoplaneta é qualquer planeta que esteja fora do nosso sistema estelar. Em sua maioria, um planeta que orbita uma ou mais estrelas que não sejam o Sol [9]. Planetas que não possuem uma estrela hospedeira e vagam sozinhos através da galáxia também podem ser considerados exoplanetas, mas neste trabalho focaremos apenas em planetas com estrela hospedeira.

Assim como no nosso sistema solar, os exoplanetas podem ser gasosos ou rochosos e terem características totalmente diferentes, seja em relação ao tamanho, massa ou composição. Ao utilizar vários métodos de detecção de exoplanetas em conjunto, é possível calcular o diâmetro e a massa desses planetas. Além de fazer estimativas da temperatura e níveis de luz [9].

Até o momento já foram confirmadas a existência de 5014 exoplanetas, e existem 8887 candidatos à espera de confirmação segundo o catálogo de exoplanetas da NASA [8].

2.2 MÉTODOS DE DETECÇÃO

Assim como listado em [17], existem diversos métodos para detectar exoplanetas. Sendo eles os métodos de Trânsito, Velocidade Radial, Imagem direta, Astrometria e Microlente Gravitacional. Dentre esses métodos, aqueles que mais possuem exoplanetas identificados são os de Trânsito, Velocidade Radial, Microlente Gravitacional e Imagem direta. O método de Astrometria só conseguiu identificar um exoplaneta até o momento, segundo o catálogo de exoplanetas da NASA [8]. Nesta seção, explicaremos como funcionam alguns desses métodos, e quais foram suas contribuições.

2.2.1 VELOCIDADE RADIAL

O método utilizado para a descoberta do primeiro exoplaneta foi o de Velocidade Radial [2], que seguiu sendo o principal método até o fim do século XX [6]. Este método

consiste na observação do efeito da gravidade que um planeta tem em sua estrela hospedeira, pois uma estrela sem planetas tem seu centro de gravidade em seu centro. Por outro lado, uma estrela com um sistema planetário terá seu centro de gravidade alterado, pois a gravidade dos planetas também influencia na estrela. Por isso, a estrela se movimenta para frente e para trás na nossa perspectiva aqui na terra, causando o efeito doppler nas suas ondas de luz [5].

Porém, planetas relativamente pequenos, com tamanho parecido com o da Terra e Marte tem pouca influência na gravidade de sua estrela. Por isso, planetas maiores como os gigantes gasosos são mais facilmente percebidos e detectados por esse método [5].

Este método conseguiu encontrar até hoje 919 exoplanetas, sendo 844 gigantes gasosos parecidos com Júpiter ou Saturno e apenas 1 planeta rochoso parecido com a Terra [8]. O método de Velocidade radial também é muito utilizado para confirmação de exoplanetas inicialmente identificados por outros métodos servindo como um método de prova [5]

2.2.2 TRÂNSITO

O método de trânsito consiste na procura pela diminuição da emissão de luz de uma estrela, pois quando um planeta passa entre um observador e a estrela, enquanto realiza sua órbita, ocorre um escurecimento momentâneo da estrela, assim como em um eclipse solar que nós conhecemos [10]. Portanto, este método analisa dados de curvas de luz que são referentes a emissão de luz de uma estrela através do tempo, como podemos ver na figura 1.

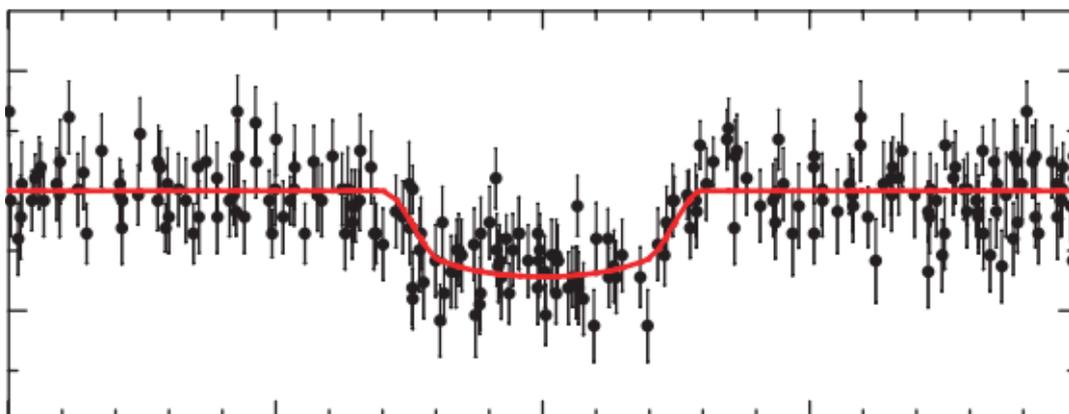


Figura 1: Curva de Luz do primeiro exoplaneta descoberto pelo método de trânsito

Fonte: Konacki, M., Torres, G., Jha, S. et al. (2003)

O primeiro exoplaneta encontrado pelo método de trânsito foi o OGLE-TR-56 b, que foi descoberto por Maciej Konacki em 2003 [10]. Ao contrário do que vimos no método de Velocidade Radial, em que só foi encontrado 1 exoplaneta parecido com a Terra, o método de Trânsito consegue detectar mais facilmente planetas parecidos com a Terra. Por isso, o número de exoplanetas rochosos e com tamanho aproximado ao da Terra encontrados pelo método de Trânsito são, até hoje, 182. No total, o número de exoplanetas encontrados por este método é 3846, sendo atualmente o método que mais detectou exoplanetas na história [8]

2.2.3 IMAGEM DIRETA

O método de imagem direta foi utilizado pela primeira vez para detectar um exoplaneta em 2004 [6] e conseguiu encontrar até hoje 59 exoplanetas [8]. Este método se baseia na visualização direta de um exoplaneta por meio do bloqueio da luz emitida pela estrela. Ao bloquear a luz de uma estrela por meio de instrumentos no telescópio, se existir um planeta orbitando a estrela, e este planeta for grande o suficiente e estiver longe o suficiente para também não ser bloqueado, será possível capturar imagens desses exoplanetas [5]. Um exemplo de um planeta visualizado por imagem direta pode ser visto na Figura 2.

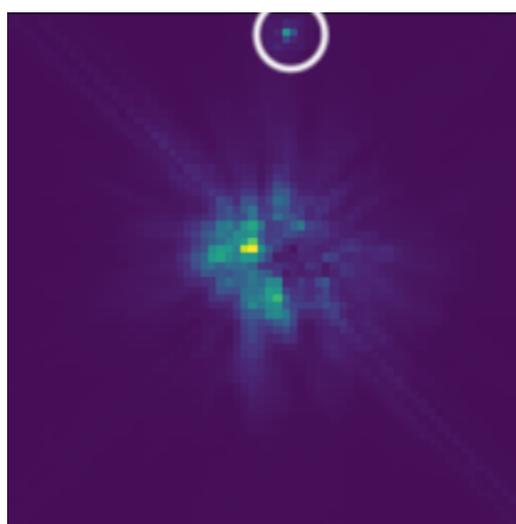


Figura 2: Exemplo de Imagem Direta de um exoplaneta

Fonte: Hou Yip, Kai (2019)

2.3 INTELIGÊNCIA ARTIFICIAL

Michael Haenlein e Andreas Kaplan [11] concluem que Inteligência Artificial (IA) se refere à habilidade de máquinas ou sistemas de interpretar dados, aprender com eles, e conseguir usar o aprendizado para atingir objetivos e resolver problemas. E para completar essa definição, Dimiter Dobrev [12] define que IA resolve tais problemas de forma igual ou superior a um humano, fazendo jus, assim, a seu nome. Ou seja, IA simula a inteligência humana de forma artificial, ao fazer com que computadores, máquinas ou sistemas possam agir ou realizar atividades de forma inteligente.

3 METODOLOGIA

Este capítulo tem como objetivo apresentar e especificar as etapas que serão seguidas para o cumprimento da metodologia que será utilizada para a execução deste trabalho.

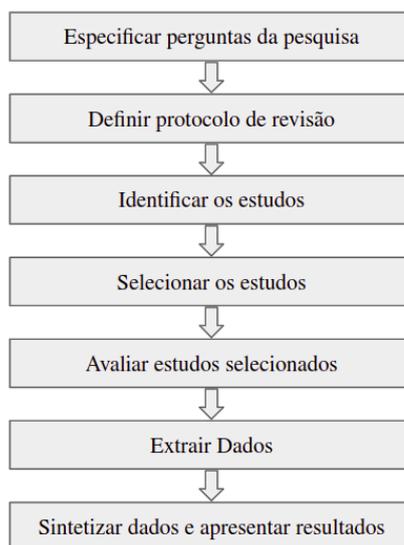
3.1 REVISÃO SISTEMÁTICA DE LITERATURA

A metodologia utilizada neste trabalho foi uma Revisão Sistemática da Literatura. Esta metodologia agrega resultados de pesquisas primárias e fornece um resumo objetivo do tema em questão, como dito por Brereton em [13]. Já Xiao em [14] diz que este método de pesquisa tem grande importância no avanço do conhecimento, pois com ele, é possível compreender qual o cenário existente e identificar possíveis lacunas que podem ser exploradas.

3.2 ETAPAS DA REVISÃO

Para a execução deste trabalho iremos nos basear nas etapas propostas por Kitchenham em [15]. A figura abaixo apresenta as etapas para a realização da revisão sistemática de literatura.

Figura 3: Detalhamento das etapas do método proposto por Kitchenham



Fonte: Adaptado de Kitchenham (2007)

O primeiro passo, e o coração de uma Revisão Sistemática de Literatura, é aquele em que especificamos a pergunta da pesquisa. Esta pergunta será a questão chave que o trabalho deseja responder. Kitchenham [15] também propõe a especificação de perguntas secundárias, para que seja possível aprofundar ainda mais a questão da pesquisa.

Kitchenham também propõe a definição de um protocolo da revisão que funciona como um planejamento, para o que será utilizado nas etapas de execução. No protocolo da revisão definiremos as questões da pesquisa, as palavras chaves, string de busca, fontes de busca, os critérios de inclusão e exclusão, que serão regras que deverão ser cumpridas para que um trabalho seja aprovado e considerado como base para esta revisão, procedimentos de seleção dos estudos, estratégia de avaliação de qualidade e de extração de dados.

A etapa de Identificação dos estudos, que dá início a fase de execução, é a etapa na qual os trabalhos primários serão identificados utilizando as estratégias de busca.

Após a identificação, temos a etapa da seleção dos estudos, onde os estudos identificados são selecionados a partir dos critérios de inclusão e exclusão que foram definidos no protocolo de revisão.

Em seguida, os trabalhos selecionados passam por uma avaliação de qualidade para apoiar os critérios de inclusão e exclusão e assegurar a qualidade dos trabalhos primários selecionados [15].

Os trabalhos que forem aprovados entrarão na etapa da extração de dados, que é o momento onde são coletados os dados que serão úteis para responder às questões da pesquisa [15]. Então os dados extraídos são sintetizados de forma descritiva, para responder às perguntas da pesquisa e apresentados, nesta que é a última etapa de uma revisão sistemática de literatura [15].

4 PROTOCOLO DA REVISÃO SISTEMÁTICA DA LITERATURA

Esse capítulo apresenta o protocolo da revisão sistemática de literatura. Este protocolo funciona como um planejamento inicial da pesquisa, e apresentará os métodos que foram utilizados para realizá-la.

4.1 PERGUNTAS DA PESQUISA

Esta seção tem como objetivo definir as perguntas da pesquisa. Primeiramente a questão central, que já foi apresentada no início deste trabalho, e por fim as questões secundárias.

A questão central da pesquisa é: “Quais as características e contribuições das técnicas de inteligência artificial que contribuem para a identificação de exoplanetas?”

As questões secundárias são as seguintes:

- QS1 - Quais técnicas inteligentes são utilizadas atualmente para identificar exoplanetas?
- QS2 - Quais métodos de identificação de exoplanetas tiveram maior contribuição do uso de Inteligência artificial?
- QS3 - Quais técnicas apresentaram melhor desempenho?

- QS4 - Quais os principais impactos e contribuições que essas técnicas trouxeram para a identificação de exoplanetas?
- QS5 - Quais os principais desafios ou limitações que as técnicas aplicadas nos estudos enfrentaram em seus contextos?

4.2 ESTRATÉGIA DE BUSCA

4.2.1 FONTES DE BUSCA

Para termos uma maior abrangência na realização das buscas pelos estudos primários, faremos uso de diversas bibliotecas acadêmicas. Utilizaremos as fontes sugeridas por Brereton em [13], além da *NASA Astrophysics Data System (ADS)*, que é uma fonte especializada para trabalhos sobre astronomia e astrofísica, e foi criada pela NASA com o objetivo de ser a fonte oficial da astrofísica [18]. A ADS é um recurso especial que é fácil de ser utilizado tanto por um pesquisador avançado, quanto por um iniciante, segundo Robbins em [18]. A tabela 1 apresenta as bases de dados que serão utilizadas neste trabalho

Fonte	Endereço
IEEEExplore	http://ieeexplore.org
ACM Digital library	http://dl.acm.org
ScienceDirect	http://www.sciencedirect.com/
SCOPUS	http://scopus.com
ResearchGate	http://researchgate.net
ADS astrophysics data system	https://ui.adsabs.harvard.edu/

Tabela 1: Bases de dados utilizadas
Fonte: Elaborado pelo autor

4.2.2 TERMOS DE BUSCA

Para definir os termos de busca, assim como sugerido por Kitchenham [15], selecionamos palavras-chave que derivam da questão da pesquisa, como por exemplo a tecnologia e a área da ciência que estão sendo analisadas. Em seguida, com o intuito de ampliar o alcance da busca, foram identificados sinônimos para cada palavra chave. A tabela 2 apresenta as palavras-chave e seus respectivos sinônimos.

Palavras-chave	Sinônimos
Artificial Intelligence	machine learning, deep learning
Detection	identification, detecting, identify, detecting, identifying
Exoplanet	extrasolar planet

Tabela 2: Palavra-chave e seus respectivos sinônimos

Fonte: Elaborado pelo autor

4.2.3 STRING DE BUSCA

Para realizar as buscas é necessário definir uma string de busca, onde são utilizados conectores lógicos. O operador “AND” é usado para agregar as palavras chaves e o operador “OR” para incluir os seus respectivos sinônimos [15]. A tabela 3 apresenta a string de busca que será utilizada neste trabalho:

<i>String de busca</i>
<i>("artificial intelligence" OR "machine learning" OR "deep learning") AND ("detection" OR "identification" OR "Detect" OR "identify" OR "detecting" OR "identifying") AND ("exoplanet" OR "extrasolar planet")</i>

Tabela 3: *String* de busca da pesquisa

Fonte: Elaborado pelo autor

4.3 CRITÉRIOS DE SELEÇÃO DOS ESTUDOS

Os critérios de inclusão e exclusão são utilizados para reduzir a probabilidade de viés e selecionar os estudos capazes de responder às questões da pesquisa, e podem ser ajustados no decorrer da pesquisa [15]. Os critérios de exclusão (CE) e critérios de inclusão (CI) são definidos nas tabelas 4 e 5, respectivamente.

4.3.1 CRITÉRIOS DE EXCLUSÃO

Critério	Descrição
CE1	Publicações que não estejam disponíveis para download ou visualização de forma gratuita mesmo com acesso institucional.
CE2	Trabalhos que foram publicados antes do ano de 2018
CE3	Publicações que não sejam estudos primários
CE4	Publicações duplicadas
CE5	Publicações que não tratam do tema "Inteligência artificial para identificação de exoplanetas"

Tabela 4: Critérios de exclusão
Fonte: Elaborado pelo autor

4.3.2 CRITÉRIOS DE INCLUSÃO

Critério	Descrição
CI1	Publicações disponíveis para download ou visualização
CI2	Trabalhos que foram publicados no período de 2018 até 2022
CI3	Publicações que sejam estudos primários
CI4	Publicações que tratam do tema "Inteligência artificial para identificação de exoplanetas"

Tabela 5: Critérios de inclusão
Fonte: Elaborado pelo autor

4.4 PROCEDIMENTOS DA SELEÇÃO DOS ESTUDOS

Os critérios de seleção definidos em 4.3 foram aplicados nos estudos resultantes das buscas realizadas de acordo com a estratégia de busca definida em 4.2. Os critérios foram aplicados em ordem e de forma sequencial para reduzir retrabalho. Para garantir que os critérios CI4 e CE5 fossem avaliados precisamente, foi feita a leitura do título e do resumo, e se ainda restou dúvidas, da introdução e conclusão dos artigos que foram incluídos nos demais critérios de inclusão, e não foram excluídos durante a aplicação dos outros critérios de exclusão. Por isso estes foram os últimos critérios a serem aplicados.

4.5 CRITÉRIOS DA AVALIAÇÃO DE QUALIDADE

A partir desta etapa os estudos restantes receberam um ID com o objetivo de referenciá-los e identificarmos de forma mais fácil. Para avaliar os estudos que passaram pelos critérios de seleção, Kitchenham [15] propõe uma avaliação de qualidade dos estudos para que tenhamos critérios de exclusão ainda mais detalhados. Para isso, foi feita a leitura completa de cada um dos estudos primários restantes, que passaram por uma avaliação dos seguintes critérios que foram adaptados de Kitchenham [15]:

CQ1 - Quão bem definida foi a Metodologia da pesquisa?

CQ2 - Quão relevantes são os resultados?

CQ3 - Quão bem definido foi o contexto?

CQ4 - Quão bem a pesquisa conseguiu atingir seus objetivos iniciais?

CQ5 - Quão claro está a apresentação dos resultados?

Para quantificar a avaliação, utilizamos as pontuações utilizadas por Kitchenham em [16]: 0, se não atender ao critério, 0,5 se atender parcialmente ao critério avaliado. E 1 se atender amplamente ao critério avaliado. foi calculada a média de cada trabalho para os 5 critérios de avaliação de qualidade e de acordo com o resultado, os trabalhos foram incluídos ou excluídos conforme a tabela abaixo onde N é o valor da média obtida:

Classificação de acordo com a nota N	
$0 \leq N < 0.7$	Excluídos
$0.7 \leq N \leq 1$	Incluídos

Tabela 6: Classificação dos estudos de acordo com a média obtida
Fonte: Elaborado pelo autor

4.6 ESTRATÉGIA DA EXTRAÇÃO DOS DADOS

A extração dos dados ocorreu em duas etapas. A primeira foi feita para a identificação dos estudos, onde foram armazenados o título, autor e a base de busca em que o trabalho foi encontrado, assim como proposto por Kitchenham [15]. Na segunda etapa foram coletadas informações que trazem uma visão geral dos estudos, como as técnicas que os estudos utilizam, e qual o método de detecção de exoplanetas que o estudo utiliza. Além disso, é nessa etapa que foram coletadas as informações necessárias para responder às questões da pesquisa.

5 SELEÇÃO DOS ESTUDOS E APRESENTAÇÃO DOS RESULTADOS

5.1 IDENTIFICAÇÃO DOS ESTUDOS

Para identificar os estudos, foram realizadas as buscas utilizando a estratégia definida no protocolo de revisão, e para cada trabalho resultante das buscas, armazenamos o título, autor, ano de publicação e a base de busca em que o trabalho foi encontrado, em uma planilha do Google. Foram identificados 427 estudos no total. O resultado detalhado da identificação dos estudos pode ser visto na figura abaixo:

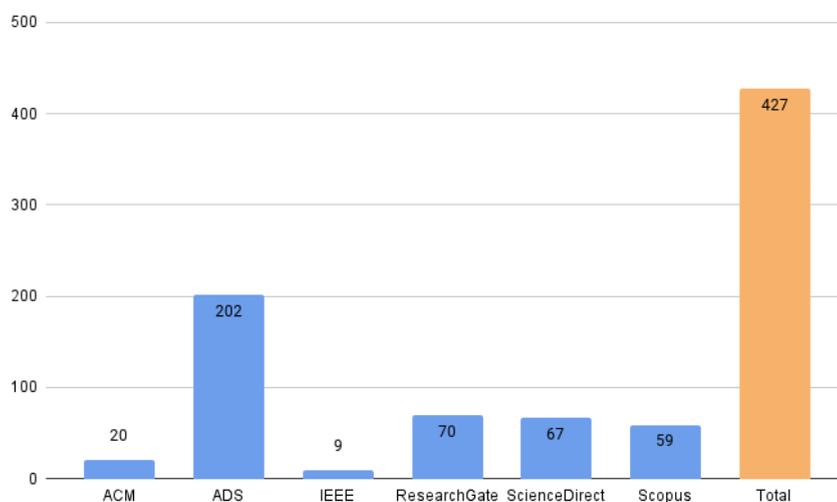


Figura 4: Estudos identificados em cada base de busca

Fonte: Elaborado pelo autor

5.2 SELEÇÃO DOS ESTUDOS

Nesta etapa foram aplicados os critérios de exclusão definidos em 4.3 para selecionar os estudos que foram utilizados nas etapas seguintes. Os critérios foram aplicados de acordo com a ordem que foram definidos e de forma sequencial. Desta forma, se um artigo for excluído na aplicação do critério CE1, não será necessário aplicar os demais critérios para este estudo. Na tabela abaixo é possível ver a quantidade de estudos que foram excluídos após a aplicação de cada critério de exclusão.

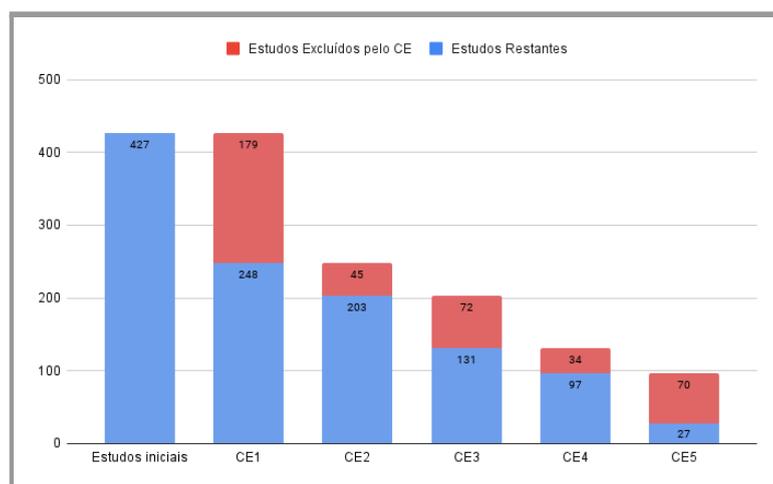


Figura 5: Estudos selecionados após aplicação de cada critério

Fonte: Elaborado pelo autor

Como mostrado na figura, dos 427 estudos identificados, 179 foram excluídos devido a não estarem disponíveis para download ou visualização gratuita. Após isso, dos 203 estudos restantes, 45 foram excluídos por se tratarem de trabalhos que foram publicados antes do ano de 2018. Dentre os testes que restaram da aplicação dos critérios CE1 e CE2, 72 estudos não eram estudos primários e também foram excluídos. Nos 131 testes restantes, 34 eram estudos duplicados. Por fim, mais 70 estudos foram excluídos por não se tratarem exatamente do tema “Inteligência artificial para detecção de exoplanetas”, mas sim, em muitos casos, sobre a habitabilidade de exoplanetas, detecção de outros objetos espaciais e busca de tecno assinaturas em exoplanetas.

Ao final da aplicação de todos os critérios de exclusão, restaram 27 artigos. O detalhamento dos estudos identificados e estudos selecionados (ES) em cada base de busca se encontra na figura 6:

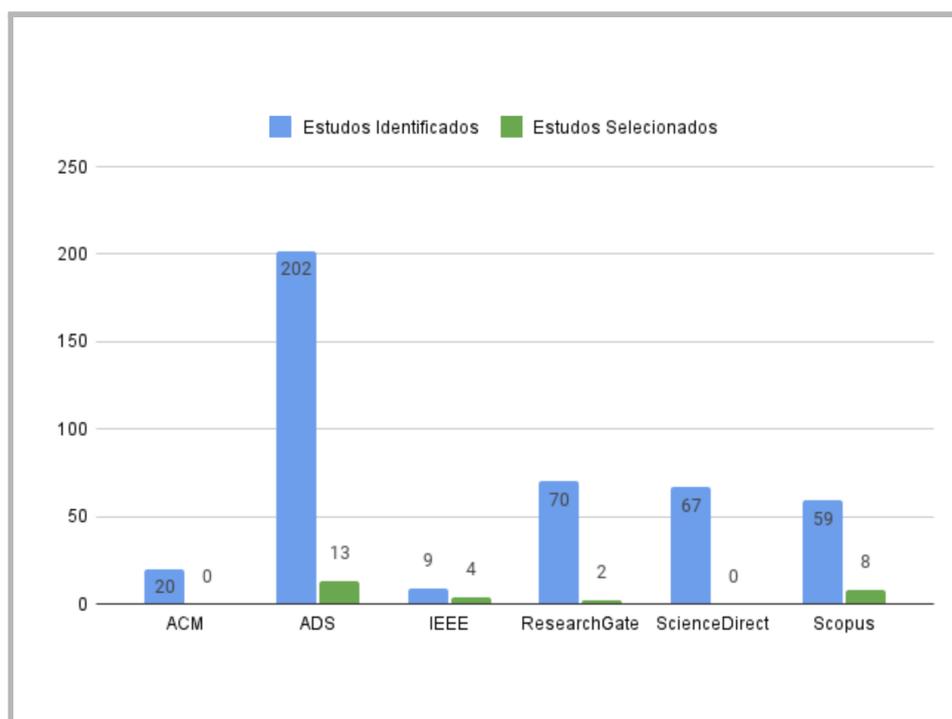


Figura 6: Estudos selecionados em cada base

Fonte: Elaborado pelo autor

5.3 AVALIAÇÃO DE QUALIDADE

Assim como definido em 4.4, foram definidos IDs para cada estudo restante. Esses ID representaram seus respectivos estudos até o fim deste trabalho. A tabela 5 mostra os IDs de cada estudo.

ID	Título do Estudo	Autores	Ano	Base
ES1	Searching for Possible Exoplanet Transits from BRITE Data through a Machine Learning Technique	Yeh, Li-Chin; Jiang, Ing-Guey;	2021	ADS
ES2	Detecting Exoplanet Transits through Machine-learning Techniques with Convolutional Neural Networks	Chintarungruangchai, Pattan; Jiang, Ing-Guey	2019	ADS
ES3	Pushing the Limits of Exoplanet Discovery via Direct Imaging with Deep Learning	Hou Yip, Kai; Nikolaou, Nikolaos; Coronica, Piero; Tsiaras, Angelos; Edwards, Billy; Changeat, Quentin; Morvan, Mario; Biller, Beth; Hinkley, Sasha; Salmond, Jeffrey; Archer, Matthew; Sumption, Paul; Choquet, Elodie; Soummer, Remi; Pueyo,	2019	ADS

		Laurent; Waldmann, Ingo P.		
ES4	Machine-learning approaches to exoplanet transit detection and candidate validation in wide-field ground-based surveys	Schanche, N; Collier Cameron, A; Hébrard, G.; Nielsen, L; Triaud, A. H. M. J.; Almenara, J. M.; Alsubai, K. A. ; Anderson, D. R.; Armstrong, D. J.; Barros, S. C. C.; Bouchy, F.; Boumis, P.; Brown, D. J. A.; Faedi, F.; Hay, K.; Hebb, L.; Kiefer, F.; Mancini, L.; Maxted, P. F. L.; Palte, E.	2019	ADS
ES5	TSARDI: a Machine Learning data rejection algorithm for transiting exoplanet light curves	Mislis, D; Pyrzas, S; Alsubai, K. A.	2018	ADS
ES6	Analysis of Kepler Objects of Interest using Machine Learning for Exoplanet Identification	Ameya Rajendra Bhamare; Aditeya Baral; Saarthak Agarwal.	2021	IEEE
ES7	Refining Exoplanet Detection Using Supervised Learning and Feature Engineering	Margarita Bugueño; Francisco Mena; Mauricio Araya	2018	IEEE
ES8	Anomalous Behavior Detection in Galaxies and Exoplanets using ML & DL Techniques	Shivani Dere; Maziya Fatima; Rutuja Jagtap; Unzela Inamdar; Nikhilkumar Shardoor	2021	IEEE
ES9	Detection of Exoplanet Systems In Kepler Light Curves Using Adaptive Neuro-Fuzzy System	R. M. Asif Amin; Abu Talha Khan; Zareen Tasnim Raisa; Nawar Chisty; Sumayra SamihaKhan.	2018	IEEE
ES10	A convolutional neural network (CNN) based ensemble model for exoplanet detection	Ishaani Priyadarshini Vikram Puri	2021	ResearchGate
ES11	Automatic vetting of planet candidates from ground based surveys: Machine learning with NGTS	D. J. Armstrong Maximilian N Günther Maximilian N; Günther James; McCormac Show; Peter J. Wheatley.	2018	ResearchGate
ES12	Exoplanet validation with machine learning: 50 new validated Kepler planets	Armstrong D.J., Gamper J.c, d, Damoulas T.d, e, f	2021	SCOPUS
ES13	Nigraha: Machine-learning-based pipeline to identify and evaluate planet candidates from TESS	Rao S.; Mahabal A., Rao N.d, Raghavendra C.	2021	SCOPUS
ES14	Harnessing the power of CNNs for unevenly-sampled light-curves using Markov Transition Field	Bugueño M; G.a, Mena F. Olivares P.b, Araya M.	2021	SCOPUS
ES15	Rapid classification of TESS planet candidates with convolutional neural networks	Osborn H.P., Ansdell M.b, Ioannou Y.c, Sasdelli M.d, Angerhausen D.e, f, Caldwell D.g, h, Jenkins J.M.g, Räissi C.i, Smith J.C.g, h	2020	SCOPUS

ES16	Identifying Exoplanets with Deep Learning. II. Two New Super-Earths Uncovered by a Neural Network in K2 Data	Dattilo A.a,Send mail to Dattilo A.,Vanderburg A.a,Shallue C.J.b,Mayo A.W.c,Berlind P.d,Bieryla A.d,Calkins M.L.d,Esquerdo G.A.d,Everett M.E.e,Howell S.B.f,Latham D.W.d,Scott N.J.f	2019	SCOPUS
ES17	Supervised detection of exoplanets in high-contrast imaging sequences	Gomez Gonzalez C.A.a, b,Absil O.a, Van Droogenbroeck M.c	2018	SCOPUS
ES18	Identifying Exoplanets with Deep Learning: A Five-planet Resonant Chain around Kepler-80 and an Eighth Planet around Kepler-90	Shallue C.J. Vanderburg A.	2018	SCOPUS
ES19	Identifying Exoplanets with Machine Learning Methods: A Preliminary Study	Jin, Yucheng; Yang, Lanyi; Chiang, Chia-En	2022	ADS
ES20	Identify Light-curve Signals with Deep Learning Based Object Detection Algorithm. I. Transit Detection	Cui, Kaiming search by orcid ; Liu, Junjie ; Feng, Fabo search by orcid ; Liu, Jifeng	2022	ADS
ES21	Exoplanet detection using machine learning	Malik, Abhishek ; Moster, Benjamin P. ; Obermeier, Christian	2021	ADS
ES22	Computer vision as a powerful tool for transiting exoplanet search	Klagyivik, Peter	2021	ADS
ES23	Identifying Exoplanets with Deep Learning. IV. Removing Stellar Activity Signals from Radial Velocity Measurements Using Neural Networks	de Beurs, Zoe L. ; Vanderburg, Andrew; Shallue, Christopher J. ; Dumusque, Xavier ; Collier Cameron, Andrew; Buchhave, Lars A.; Cosentino, Rosario; Ghedina, Adriano; Haywood, Raphaëlle D.; Langellier, Nicholas ; Latham, David W. ; López-Morales, Mercedes ; Mayor, Michel ; Micela, Giusi ; Milbourne, Timothy W. ; Mortier, Annelies ; Molinari, Emilio ; Pepe, Francesco ; Phillips, David F. ; Pinamonti, Matteo; Piotto, Giampaolo ; Rice, Ken ; Sasselov, Dimitar ; Sozzetti, Alessandro; Udry, Stéphane ; Watson, Christopher A.	2020	ADS
ES24	Identifying Planetary Transit Candidates in TESS Full-frame Image Light Curves via Convolutional Neural Networks	Olmschenk, Greg; Ishitani Silva, Stela; Rau, Gioia; Barry, Richard K.; Kruse, Ethan; Cacciapuoti, Luca ; Kostov, Veselin; Powell, Brian P.; Wyrwas, Edward ; Schnittman, Jeremy D.; Barclay, Thomas	2021	ADS
ES25	A one-armed CNN for exoplanet detection from light curves	Visser, Koko ; Bosma, Bas ; Postma, Eric	2021	ADS

ES15	1	1	1	0,5	1	0,9	Incluído
ES16	1	1	1	1	1	1	Incluído
ES17	1	1	1	1	0,5	0,9	Incluído
ES18	1	1	1	1	1	1	Incluído
ES19	1	1	0,5	0,5	0,5	0,7	Incluído
ES20	1	1	1	1	0,5	0,9	Incluído
ES21	1	1	0,5	1	1	0,9	Incluído
ES22	0,5	0,5	0	0,5	0	0,3	Excluído
ES23	1	1	1	1	1	1	Incluído
ES24	1	1	1	1	0,5	0,9	Incluído
ES25	1	1	1	1	1	1	Incluído
ES26	1	1	1	1	0,5	0,9	Incluído
ES27	1	1	0,5	1	0,5	0,8	Incluído

Tabela 8: Resultados da avaliação de qualidade

Fonte: Elaborado pelo autor

Podemos ver que dos 26 estudos selecionados, 24 passaram na avaliação de qualidade e apenas os artigos ES8 e ES22 foram reprovados.

5.4 EXTRAÇÃO DOS DADOS

Os dados que foram retirados de cada estudo para responder às perguntas da pesquisa foram armazenados em uma planilha. Para a Q1, foram armazenados quais as técnicas de inteligência artificial foram utilizadas para a execução do estudo e o trecho que explicita essa informação. Em alguns casos, principalmente aqueles em que são usadas mais de uma técnica, os trechos que comprovam a utilização de cada uma das técnicas não estão na mesma frase ou parágrafo. Por isso, nesses casos, foram armazenados mais de um trecho para responder esta pergunta.

Em relação a Q2, armazenamos o trecho que mostra quais métodos de detecção de exoplanetas recebem a contribuição do uso de inteligência artificial no estudo. Em alguns casos, não foi dito diretamente, por texto, qual o método de detecção foi utilizado, porém inferimos quando possível, como por exemplo no caso em que foi dito que foram utilizadas dados de curvas de luz, e apenas com essa informação, podemos afirmar que o método utilizado foi o de trânsito. Nesses casos, armazenamos o trecho em que é dito explicitamente ou não, qual o método utilizado.

Para responder a Q3, armazenamos o trecho em que é citado a acurácia, eficiência ou outro valor que demonstre o desempenho da técnica proposta. Em alguns casos, os autores informaram o valor do desempenho apenas dentro de uma tabela, por isso, o valor do desempenho não está presente no texto. Por isso, extraímos o valor da tabela, e armazenamos, além disso, o trecho em que é feita a referência à tabela em que se encontram os valores de desempenho.

Por fim, para responder a Q4 e Q5 foi feita uma análise crítica dos estudos e foram identificadas os principais diferenciais, impactos ou contribuições que seus estudos trouxeram para a área do descobrimento de exoplanetas, e quais as limitações e desafios enfrentados pelo uso IA.

5.5 APRESENTAÇÃO E SÍNTESE DOS RESULTADOS

Nesta seção, serão apresentados os resultados desta pesquisa, por meio da síntese dos dados armazenados na etapa anterior e apresentação das respostas para as perguntas da pesquisa

5.5.1 Q1: QUAIS TÉCNICAS INTELIGENTES SÃO UTILIZADAS ATUALMENTE PARA IDENTIFICAR EXOPLANETAS?

Na tabela 9, é possível ver as técnicas de inteligência artificial utilizadas nos estudos selecionados.

Técnica	Estudos
Convolutional Neural Network	ES1, ES2, ES3, ES4, ES10, ES13, ES15, ES14, ES16, ES17, ES18, ES20, ES23, ES24, ES25, ES26
Random Forest	ES4, ES6, ES7, ES10, ES11, ES12, ES17, ES19
Support-vector machine	ES6, ES7, ES10, ES26, ES27
Multilayer perceptron	ES10, ES12, ES19, ES26
Logistic regression	ES4, ES7, ES10
K-nearest neighbors	ES4, ES7
Linear Support Vector Classifier	ES4, ES23
Fully Connected Neural Network	ES18, ES23
AdaBoost	ES6
Feedforward neural network	ES6
Adaptive neuro fuzzy inference system	ES9
Decision tree	ES10
Self-organizing map	ES11
Gaussian process classifier	ES12
Extra trees	ES12
Naïve Bayes	ES19

K-means	ES19
Gradient Boosted Trees	ES21
Box fitting method	ES26
Autoencoder	ES27
Generative adversarial network	ES3
Support Vector Classifier	ES4
DBSCAN	ES5
Linear Neural Network	ES18
Wavelet MLP	ES26

Tabela 9: Técnicas utilizadas nos estudos
 Fonte: Elaborado pelo autor

Vimos que a Convolutional neural network (CNN) foi a técnica mais utilizada nos estudos selecionados. CNN é uma técnica de redes neurais que se diferencia pelo seu diferencial na área de processamento e análise de imagens [19]. As citações abaixo foram armazenadas na etapa de extração de dados e comprovam a utilização de cada técnica em seu respectivo estudo selecionado.

“In this study, we have considered the one-dimension convolutional neural network (1DCNN) of machine learning techniques. With this deep-learning technique, we decided to examine the possible transits with periods from one to five days.” - ES1

“ (...) both 1D-CNN and 2D-CNN deep learning models with phase folding will be constructed (...) five deep learning models are constructed based on convolutional neural network (CNN)” - ES2

“ We use a Generative Adversarial Network to obtain a suitable dataset for training Convolutional NeuralNetwork classifiers to detect and locate planets across a wide range of SNR[...] We use this dataset to train a Convolutional Neural Network (CNN) to classify images as positive/negative.” - ES3

“We use a combination of machine learning methods including Random Forest Classifiers (RFCs) and Convolutional Neural Networks (CNNs) to distinguish between the different types of signals [...] The training dataset is used as an input to a variety of classifiers, namely a Support Vector Classifier (...) Linear Support Vector Classifier (...), Logistic regression (...), K-nearest neighbors (...)” - ES4

“TSARDI is based on the Machine Learning clustering algorithm DBSCAN, and its purpose is to serve as a robust and adaptable filter aiming to identify unwanted noise points left over from data detrending processes. TSARDI is an unsupervised method, which can treat each light curve individually” - ES5

“For our analysis, we have used four models. Namely, Support Vector Machines (SVM), Random Forest, AdaBoost and a Feed-Forward Neural Network.” - ES6

“The second model is the regularized logistic regression [...] The K-nearest neighbors model (k-NN) is a popular yet simple approach based on memory [...] We combined the identification of Confirmed objects (exoplanets) using Random Forest and the identification of False Positive objects (not exoplanet) using a SVM RBF model to produce our classification” - ES7

“Then, we use the extracted features as inputs to an Adaptive Neuro-Fuzzy Model (ANFIS) that is trained on a set of light curves labeled by human experts” - ES9

“The techniques used are Decision Trees, Support Vector Machines, Logistic Regression, Convolutional Neural Networks (CNN), Random Forest Classifier, and Multilayer Perceptron (MLP).” - ES10

“We draw established techniques from the machine learning field, utilizing random forests (Breiman 2001) and self organizing-maps (Kohonen 1982).” - ES11

“We proceed with the RFC as a versatile robust algorithm, supplementing the results with classifications from the next two most successful models, ET and MLP, to guard against overfitting by any one model (...) we utilize a GPC to provide an independent and naturally probabilistic method for comparison and to guard against overconfidence in model classifications” - ES12

“In this paper, we explore the use of CNNs for classifying TESS light curves to search for planet candidates” - ES13

“the key idea of this article is to harness the power of Deep Learning for image processing, specifically 2D convolutional neural network” - ES14

“We adapted an existing neural network model and then trained and tested this updated model on four sectors of high-fidelity, pixel-level TESS simulations data created using the Lilith simulator and processed using the full TESS pipeline [...] We maintained the convolutional filter sizes and architecture from Astronet, with four 1D convolutional layers for the local view, and eight for the global view” - ES15

“We train a convolutional neural network, called AstroNet-K2, to predict whether a given possible exoplanet signal is really caused by an exoplanet or a false positive” - ES16

“In this new framework, we present algorithmic solutions using two different discriminative models: SODIRF (random forests) and SODINN (neural networks).” - ES17

“In this paper, we present a deep neural network for automatically vetting Kepler TCEs. [...] We consider three types of neural network for classifying Kepler TCEs as either ‘planets’ or ‘not planets.’” - ES18

“For the supervised learning part, we used decision tree, random forest, naïve Bayes, and neural network. For the unsupervised learning part, we used k-means clustering. [...] We implemented the neural network model using Multilayer Perceptron Regressor from scikit-learn library” - ES19

“In this work, we explore a two-dimensional (2D) object detection algorithm to identify transiting signals” - ES20

“Our method consists of automatically extracting time series features from light curves which are then used as input to a gradient boosted trees model.” - ES21

“(...) we trained three different ML architectures: a linear regression model, a fully connected neural network, and a convolutional neural network.” - ES23

“We present a convolutional neural network, which we train to identify planetary transit signals and dismiss false positives.” - ES24

“(...) we propose Genesis, a one-armed simplified Convolutional Neural Network (CNN) for exoplanet detection, and compare it to the more complex, two-armed CNN called Astronet.” - ES25

“In this paper we design various deep learning algorithms to recognize planetary transit features from a training data set (...) We compare the classification of different transit detection algorithms; a fully connected neural network (MLP), convolutional neural network (CNN), fully connected network with wavelet transformed input (Wavelet MLP), a support vector machine (SVM) and a box fitting Least Squares method (BLS).” - ES26

“The goal is to use the AE as a denoising autoencoder. It allows us to encode our sample into the latent space and then reconstruct it. (...) This is followed by an auto tuned SVM classifier which segregates between the two classes based on the labels provided.” - ES27

5.5.2 Q2: QUAIS MÉTODOS DE IDENTIFICAÇÃO DE EXOPLANETAS TIVERAM MAIOR CONTRIBUIÇÃO DO USO DE INTELIGÊNCIA ARTIFICIAL?

Método de identificação	Estudos
Trânsito	ES1, ES2, ES4, ES5, ES6, ES7, ES9, ES10, ES11, ES12, ES13, ES14, ES15, ES16, ES18, ES19, ES20, ES21, ES24, ES25, ES26, ES27
Imagem direta	ES3, ES17
Velocidade Radial	ES23

Tabela 10: Métodos utilizados para detecção
Fonte: Elaborado pelo autor

“All five versions of these CNN models were employed to search for possible transits” - ES1

“A machine learning technique with two-dimension convolutional neural network is proposed for detecting exoplanet transits” - ES2

“We have demonstrated that deep learning offers a promising direction for taking the field of direct imaging of exoplanets to the next step” - ES3

“ The machine-learning framework we have created provides a tool for the observer wanting to re-examine the full set of data holdings in any WASP field, enabling fast re-classification of all targets showing transit-like behavior and identification of new targets of interest” - ES4

“The major test for our algorithm was to investigate whether it can indeed (positively) affect the transit detection efficiency” - ES5

“In contrast, Kepler’s transit time data allows for the detection of smaller Earth-sized exoplanets. This has opened up a new window of planets to be discovered.” - ES6

“Considering that most of the discovered exoplanets have been detected through the transit method, and taking advantage of the photometric improvements of Kepler, we propose to work with the Kepler Objects of Interest (KOI4) dataset” - ES7

“The paper investigates the usability of fuzzy inference systems, namely ANFIS to assist in the detection of exoplanet transit events from time-series data of Flux intensify values from host stars” - ES9

“In order to detect an exoplanet, we have considered changes influx or light intensity of the star.” - ES10

“We have presented a method for ranking candidates from a transiting planet search.” - ES11

“Our models can validate thousands of unseen candidates in seconds once applicable vetting metrics are calculated, and can be adapted to work with the active Transiting Exoplanet Survey Satellite (TESS) mission, where the large number of observed targets necessitate the use of automated algorithms” - ES12

“Nigraha uses a combination of transit finding, supervised machine learning, and detailed vetting to identify with high confidence a few planet candidates that were missed by prior searches.” - ES13

“This article focuses on identifying exoplanets only from (unfolded) light curve data, without considering other sources of information such as stellar parameters obtained through cross-matching catalogs” - ES14

“Here we created the first deep-learning model capable of classifying TESS planet candidates” - ES15

“We have built and trained a neural network to classify possible planet signals detected in K2 data into likely planet candidates and false positives [] The results of the fits are reported in Table 2 and Figure 8 shows the transit light curves with the best-fit models and the residuals.” - ES16

“two algorithms are proposed: SODIRF, which stands for Supervised exoplanet detection via Direct Imaging with Random Forests, and SODINN, which stands for Supervised exoplanet detection via Direct Imaging with deep Neural Networks” - ES17

“We present a method for classifying potential planet signals using deep learning, a class of machine learning algorithms that have recently become state-of-the art in a wide variety of tasks. We train a deep convolutional neural network to predict whether a given signal is a transiting exoplanet or a false positive caused by astrophysical or instrumental phenomena.” - ES18

“The Kepler dataset was used to predict the existence of exoplanet candidates by classification techniques” - ES19

“In this work, we explore a two-dimensional (2D) object detection algorithm to identify transiting signals” - ES20

“We introduce a new machine learning based technique to detect exoplanets using the transit method” - ES21

“Exoplanet detection with precise radial velocity (RV) observations is currently limited by spurious RV signals introduced by stellar activity. We show that machine learning techniques such as linear regression and neural networks can effectively remove the activity signals (due to starspots/faculae) from RV observations” - ES23

“We present our convolutional NN, which we train to identify planetary transit signals and dismiss false positives” - ES24

“CNNs such as Astronet and Exonet uncover exoplanets based on transit data” - ES25

“As validated by our deep net analysis of Kepler light curves, we detect periodic transits consistent with the true period without any model fitting.” - ES26

“We apply the system to TESS observations of TCE’s in search of transiting exoplanet signatures in the large TCE dataset.” - ES27

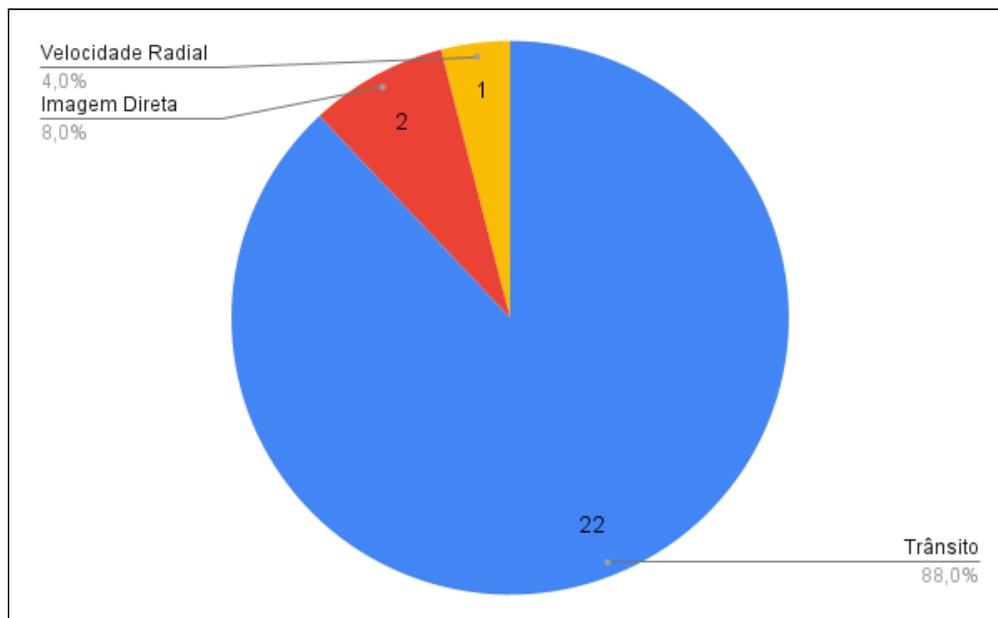


Figura 7: Utilização dos métodos de detecção

Fonte: Elaborado pelo autor

Como podemos ver na figura 7, o método de trânsito foi o método que teve mais contribuições de técnicas de inteligência artificial. Sendo seguido pelo método de imagem direta, que foi contemplado com 2 estudos selecionados que utilizaram a inteligência artificial neste método.

Na tabela 13, é possível ver as técnicas de IA e a quantidade de vezes que foram utilizadas em cada uma das contribuições e impactos identificados.

Técnicas utilizadas	Trânsito	Imagem direta	Velocidade Radial
Convolutional Neural Network	13x	2x	1x
Random Forest	7x	1x	
Multilayer perceptron	4x		
Gaussian process classifier	1x		
Extra trees	1x		

Fully Connected Neural Network	1x		1x
Linear Neural Network	1x		1x
Support-vector machine	5x		
Autoencoder	1x		
Logistic regression	3x		
K-nearest neighbors	2x		
Linear Support Vector Classifier	1x		
Support Vector Classifier	1x		
AdaBoost	1x		
Feedforward neural network	1x		
Box fitting method	1x		
Wavelet MLP	1x		
DBSCAN	1x		
Decision tree	1x		
Adaptive neuro fuzzy inference system	1x		
Self-organizing map	1x		
Naïve Bayes	1x		
K-means	1x		
Gradient Boosted Trees	1x		
Generative adversarial network		1x	

Tabela 11: Técnicas utilizadas para cada método de detecção.

Fonte: Elaborado pelo autor

5.5.3 Q3: QUAIS TÉCNICAS APRESENTARAM MELHOR DESEMPENHO?

Técnicas	Estudos
CNN	ES1, ES2, ES3, ES4, ES10, ES13, ES14, ES15, ES16, ES17, ES18, ES20, ES23, ES24, ES25
DBSCAN	ES5
AdaBoost	ES6
Random Forest	ES7, ES11
Adaptive neuro fuzzy inference system	ES9
Gaussian process classifier	ES12
Multilayer perceptron	ES19
Gradient Boosted Trees	ES21
Wavelet MLP	ES26
Support Vector Machine	ES27

Tabela 12: Técnicas com melhor desempenho
Fonte: Elaborado pelo autor

Convolutional Neural Networks

*“The **convolutional neural networks** were trained with synthetic transit signals combined with BRITE light curves until the accuracy rate was higher than 99.7 %.” - ES1*

“All models with folding can have accuracy above 98% even when SNR is less than 10. The precision and recall have a similar trend.” - ES2

“Table 3 shows the performance of the CNNs trained on the same SNR when evaluated on the 4 different test sets [...] We see that the lower the SNR on which the CNN is trained, the better its predictive performance across the entire SNR spectrum.” - ES3

“As a comparison, for our RFC, 182 objects were classified as planets, with 45 true positives and 137 false positives, indicating a success rate of 25%. The SVC is more conservative, finding fewer total planets but rejecting more false positives, and has a 49% estimated follow-up accuracy (True positives divided by true positives and false positives). The CNN with the full lightcurve showed even better results, with 81% estimated follow-up accuracy, and when the local lightcurve data was added 75% of the objects flagged as planets were true positives.” - ES4

“The Proposed Ensemble-CNN model performs relatively better with respect to the other baseline models used in the study. (...) our proposed model outperforms them all with an accuracy of 99.62%.” - ES10

“Our model has an AUC (area under the curve) score of 0.908. We also evaluate our model by augmenting our test set with bona fide planets and find that it recovers 91.9 per cent” - ES13

“(...) our method is faster to execute and lighter in terms of memory consumption, yet it offers a competitive performance of 77.01% in the F1-score (macro averaged) with respect to similar techniques. This value reflects the good behavior reached on the False Positive class (non-exoplanets) with an F1 of 84% while a 70% for the Confirmed class (true exoplanets), which is the second best result on this class” - ES14

“We achieve average precision as high as 97.3% with accuracy on planet populations as high as 91.8%.” - ES15

“Our neural network is quite successful: it is able to correctly classify signals with an accuracy of 98% and an AUC of 0.988 on our test set” - ES16

“This study shows the improved sensitivity versus specificity trade-off of the proposed supervised detection approach. At the diffraction limit, SODINN improves the true positive rate by a factor ranging from ~2 to ~10” - ES17

“Our model is highly effective at ranking individual candidates by the likelihood that they are indeed planets: 98.8% of the time it ranks plausible planet signals higher than false-positive signals in our test set” - ES18

“Through training the network on the light curves of the confirmed Kepler exoplanets, our model yields about 90% precision and recall for identifying transits with signal-to-noise ratio higher than 6 (set the confidence threshold to 0.6)” - ES20

“Our best performing models were the FC NN and CNN (Figure 6) which reduced the RV scatter (...), from 146.9 cm s⁻¹ to 78.4 cm s⁻¹ and 77.6 cm s⁻¹ respectively across the full dataset” - ES23

“Our network performs inference on a TESS 30-minute cadence light curve in ~5ms on a single GPU, enabling large scale archival searches” - ES24

“The first row of table 2 shows the performances of Genesis on the Astronet dataset, alongside the input size and number of parameters” - ES25

DBSCAN

“Taking into account the accuracy of the data used, and limiting the sample in terms of host star magnitude ($m < 12$), leads to a detection rate of 80% after using TSARDI” - ES5

AdaBoost

“However, in terms of numbers the AdaBoost classifier obtains the best performance with an average F-1 score of 98%” - ES6

Random Forest

“We reached an 88.3% of average precision and recall (f1 score), meaning that approximately the 88% of the times the prediction was correct. For this configuration the best learned model was Random Forest” - ES7

“Through incorporating random forests and self-organizing-maps we are able to obtain an AUC score of 97.6% on data from the NGTS survey” - ES11

Adaptive neuro fuzzy inference system

“The inclusion of DTW to our feature set has increased the performance significantly. The RMS error is dropped to around 0.29 we and a maximum accuracy of 81% is observed after multiple runs.”
- **ES9**

Gaussian process classifier

“The final performances of each model are given in Table 3 [...] For these cases our GPC appears to be more reliable, potentially as it is trained on the full Kepler set of FPs rather than limited to specific scenarios that may not fully explore unusual cases, or reliably account for the candidate distributions in the Kepler candidate list.” - **ES12**

Multilayer perceptron

“the decision tree achieved an accuracy of 99.06%, random forest achieved an accuracy of 92.11%, naïve Bayes achieved an accuracy of 88.50%, and multilayer perceptron achieved an accuracy of 99.79%”
- **ES19**

Gradient Boosted Trees

“For the Transiting Exoplanet Survey Satellite (TESS) data, we found our method can classify light curves with an accuracy of 0.98 (...)” - **ES21**

Wavelet MLP

“Figure 9 shows the accuracy of our deep learning algorithms on data with varying amplitudes of noise” - **ES26**

Support Vector Machine

“The effectiveness of the algorithm and its evaluation are demonstrated on Kepler data in Table 1 by identifying known planet candidate” - **ES27**

5.5.4 Q4: QUAIS OS PRINCIPAIS IMPACTOS E CONTRIBUIÇÕES QUE ESSAS TÉCNICAS TROUXERAM PARA A IDENTIFICAÇÃO DE EXOPLANETAS?

Impactos e Contribuições	Estudos
Detecção de candidatos ou confirmação de exoplanetas	ES1, ES12, ES13, ES14, ES15, ES16, ES18, ES24, ES27
Análise de estrelas mais brilhantes	ES1
Uso de modelos híbridos	ES4, ES6, ES26, ES27
Ganhos de performance	ES5, ES7, ES10, ES21

Tabela 13: Impactos e contribuições
Fonte: Elaborado pelo autor

Detecção de candidatos ou confirmação de exoplanetas

Muitos dos estudos selecionados conseguiram tanto detectar novos candidatos, quanto confirmar novos exoplanetas por meio dos modelos implementados nos estudos. O ES1, por exemplo, identificou 2 novos candidatos que serão avaliados e classificados em trabalhos futuros. Já o estudo ES12 conseguiu confirmar 50 novos exoplanetas, ao utilizar um novo algoritmo que utiliza *Gaussian process classifier*.

O estudo ES13 conseguiu encontrar 38 candidatos a exoplanetas de dados da Transiting Exoplanet Survey Satellite (TESS) por meio de seu pipeline baseado em Deep Learning, o Nigraha. Já o estudo ES14 conseguiu identificar 123 objetos com alta probabilidade de serem candidatos a exoplanetas, mas que precisam de confirmações e análises futuras, e o ES15 utilizou CNN para classificação de candidatos a exoplanetas da Transiting Exoplanet Survey Satellite (TESS) para identificar falsos positivos. E conseguiu identificar 14 falsos positivos dentre os objetos de interesse do projeto TESS.

Os estudos ES16 e ES18 identificaram 2 novos exoplanetas cada um. O ES16 identificou os exoplanetas K2-293 b e K2-294 b, e o ES18 conseguiu encontrar um planeta no

sistema Kepler-80 e o oitavo planeta do sistema Kepler-90. Com isso, esse sistema empata com o nosso sistema solar como o sistema conhecido com mais planetas confirmados.

Por fim, os estudos ES24 e ES27, conseguiram identificar 181 e 3 novos candidatos, respectivamente.

“(...) to examine the BRITTE light curves of 35 stars, 10 stars exhibited possible transits detected (...) two of them, HD37465, and HD186882 systems, were followed up through future observations with a higher priority” - ES1

“Our method has proved successful and able to validate planets rapidly.[.] We discuss the limitations and caveats of this methodology, and after accounting for possible failure modes newly validate 50 Kepler candidates as planets, sanity checking the validations by confirming them with VESPA using up to date stellar information.[.] We also introduce a GPC for exoplanet candidate vetting for the first time” - ES12

“The core contribution of this paper is in the design and evaluation of a deep learning-based pipeline for identifying planetary candidates from TESS data. In building our pipeline [...] we provide details of our pipeline, release our supervised machine learning model and code as open source, and make public the 38 candidates we have found in seven sectors” - ES13

“In this paper, we propose a method that only requires the raw light curve to identify exoplanets without the need of additional metadata or specific formats for the time series [...] Our proposed method was applied to the candidate objects that are still under study (unlabeled objects), in some cases are stars with already confirmed exoplanets. We found 96 objects with confidence probability above 95% and 27 above 99%. These could be considered as strong exoplanet candidates [...] one of the potential uses of the proposed method is to rank and identify potential candidates rather than classify them” - ES14

“Our use of multi-class models may also aid targeted follow-up observations by providing class probabilities for different false-positives which may impact follow-up strategy (...) We also identify 14 TOIs as likely false positives.” - ES15

“We identified two new exoplanets in K2 data using our averaged neural network model. Both planets are super-Earths orbiting G-dwarf stars; we list the stellar and planetary parameters for these two systems in Table 2. (...) and were able to validate these objects as genuine exoplanets. The new planets are both in close orbits around

their host stars and are intermediate to the Earth and Neptune in size”
- **ES16**

“We statistically validated two of these new candidates: Kepler80 g, which is part of an interesting five-planet resonant chain, and Kepler-90 i, which is now the eighth confirmed planet in the Kepler-90 system. This paper brings Kepler-90 into a tie with our Sun as the star known to host the most planets. [...] The immediate impact of this work is the discovery of several new Kepler planets and planet candidates, including the exciting discovery of an eighth planet in the Kepler-90 system” - **ES18**

“We present 181 new planet candidates identified by our network, which pass subsequent human vetting designed to rule out false positives. Our neural network model is additionally provided as open-source code for public use and extension.” - **ES24**

“(...) the algorithm yields about 50 targets for further analysis, and we uncover three new exoplanetary candidates by further manual vetting.” - **ES27**

Análise de estrelas mais brilhantes

Cada missão de busca por exoplanetas consegue monitorar estrelas com determinadas magnitudes. A missão Kepler por exemplo é capaz de monitorar estrelas com magnitude entre 11 e 14. Durante a análise dos estudos, foi possível perceber que apenas o estudo ES1 utilizou dados do Satélite BRITE, que foi construído para analisar estrelas de magnitude menor que 5. Com isso, esse estudo conseguiu analisar dados de estrelas mais brilhantes, que até então, não eram monitoradas por projetos de observação de estrelas como Kepler, CoRoT, HATNet e WASP, e termos dados relacionados à existência de exoplanetas em estrelas mais brilhantes do céu noturno.

“The Kepler Space Telescope can monitor stars with an apparent magnitude around $V = 11 \sim 14$. (...) CoRoT (...) $V = 12 \sim 15$. Finally, HATNet and WASP projects (...) $V = 9 \sim 12$ (...) However, the brighter stars with $V < 5$ mag remain unexplored (...) To explore the possible existence of exoplanets around brighter stars, we searched for exoplanet transits from photometrical data obtained by the BRITE satellite through a machine learning technique (...)” - **ES1**

Uso de modelos híbridos

Os estudos ES4, ES6, ES26 e ES27 mostraram que ao utilizar diversas técnicas em conjunto é possível obter resultados ainda melhores, do que se apenas escolhermos a que obtiver maior acurácia.

“Using multiple machine learning models is an effective framework that can be modified and applied to a variety of different large-scale surveys in order to reduce the total time spent in the target identification and ranking stage of exoplanet discovery. Combining the results from additional machine learning methods could further improve the predictions.” - ES4

“This paper identifies the caveats against the use of single-method planet validation techniques and cautions against it. (...) This shows that models which rank features differently can be combined together sequentially to improve overall performance.” - ES6

“(...) we test various methods to improve our planet detection rates including 1D convolutional networks and feature transformations such as wavelets and find significant improvement with CNNs. Machine learning techniques provide an artificially intelligent platform that can learn subtle features from large data sets in a more efficient manner than a human.” - ES26

“This study demonstrates for the first time the successful application of the particular combined multiple AI/ML based methodologies to a large astrophysical dataset for rapid automated classification of TCEs” - ES27

Ganhos de performance

Os estudos ES5, ES7, ES10 e ES21 deixam claro por meio de comparações e análises que seus modelos possuem um aumento de eficiência. O estudo ES7, por exemplo, mostrou os ganhos da utilização de uma abordagem supervisionada. Já o estudo ES10 fornece dados comparativos entre diversas técnicas de Inteligência artificial no campo da identificação de exoplanetas por meio de análise de curvas de luz. Além disso, propôs uma nova técnica baseada em Redes neurais que não tinha sido utilizada até o momento para esta tarefa.

“Based on the results of a search for (simulated) transits on a real data set, we demonstrate that TSARDI can lead to a substantial improvement of the transit detection efficiency; compared to light curves filtered with a straightforward sigma-clip algorithm, TSARDI-processed light curves showed an overall increase of ~10% in the number of detections.” - ES5

“The results of this work show that using only standard statistics to summarize the entire light curve has only a moderate performance compared to combining them with other manually extracted features by experts on astronomy related to the planet and its hosting star” - ES7

“In this paper, we used the intensity of light (flux) and ArtificialIntelligence techniques (...) to draw a comparison between different models for exoplanet detection. [...] Moreover, the Ensemble-CNN based approach has also been proposed for the study, which makes it the first and only one of its kind based on the state of art” - ES10

“We were able to demonstrate that with our machine learning method, we could identify light curves with planet signals more accurately as compared to BLS while significantly reducing the number of false positives” - ES21

Na tabela 14, é possível ver as técnicas de IA e a quantidade de vezes que foram utilizadas em cada uma das contribuições e impactos identificados.

Técnicas utilizadas	Detecção de candidatos ou confirmação de exoplanetas	Análise de estrelas mais brilhantes	Uso de modelos híbridos	Ganhos de performance
Convolutional Neural Network	6x	1x	2x	1x
Random Forest	1x		2x	2x
Multilayer perceptron	1x		1x	1x
Gaussian process	1x			

classifier				
Extra trees	1x			
Fully Connected Neural Network	1x			
Linear Neural Network	1x			
Support-vector machine	1x		3x	2x
Autoencoder	1x		1x	
Logistic regression			1x	2x
K-nearest neighbors			1x	1x
Linear Support Vector Classifier			1x	
Support Vector Classifier			1x	
AdaBoost			1x	
Feedforward neural network			1x	
Box fitting method			1x	
Wavelet MLP			1x	
DBSCAN				1x
Decision tree				1x
Gradient Boosted Trees				1x

Tabela 14: Técnicas utilizadas em cada contribuição e impacto identificado.

Fonte: Elaborado pelo autor

5.5.5 Q5: QUAIS OS PRINCIPAIS DESAFIOS OU LIMITAÇÕES QUE AS TÉCNICAS APLICADAS NOS ESTUDOS ENFRENTARAM EM SEUS CONTEXTOS?

Desafio ou Limitação	Estudos
Cobertura do conjunto de dados do treinamento	ES1, ES2, ES3, ES6, ES9, ES11, ES12, ES20
Identificar exoplanetas com certas características	ES12, ES13, ES24, ES26
Diferenciar Falsos-Positivos	ES14, ES16, ES18
Qualidade dos dados	ES12, ES15

Tabela 15: Limites e Desafios
Fonte: Elaborado pelo autor

Cobertura do conjunto de dados do treinamento

Os estudos ES1, ES2, ES3, ES6, ES9, ES11, ES12 e ES20 apresentaram limitações relacionadas à cobertura dos dados utilizados no treinamento de seus modelos. Ao usar dados originários de apenas uma missão espacial, os estudos ES3, ES6 e ES11, estão limitados às características que esta missão procura. Como por exemplo a magnitude da estrela ou um determinado período de trânsito. Os estudos ES1 e ES2, por exemplo, utilizam dados sintéticos, o que faz com que o modelo também não consiga analisar casos que fujam das características dos dados gerados para seu treinamento. Assim como o estudo 20, que não utilizou casos de falsos positivos em seu conjunto de dados. Tudo isso faz com que o modelo treinado trabalhe apenas com um único padrão, o que faz com que sejam descartados possíveis candidatos por apenas fugirem das características comuns do conjunto de dados que foi utilizado para seu treinamento.

“They were trained with synthetic data generated by injecting planetary transit signals into the BRITTE light curves. (...) While this study demonstrated that our CNN models could successfully identify possible transit candidates, we caution that further investigation with

additional observational data is necessary to lead to final confirmations.” - ES1

“In order to achieve that goal, we will need to generate more realistic non-transit light curves ourselves. The transits by smaller planets can then be injected and improve the training of deep learning models.” - ES2

“explore direct imaging datasets from different ground and space-based instruments, with a larger number of datapoints, more positive class examples and more observations per target. This will allow us to compare the ‘sensitivity’ afforded by each instrument” - ES3

“There is a limited amount of cumulative data available openly for public use. At present, only NASA data from their Kepler mission. Availability of more data would certainly allow for better model development. [...] The results of the model cannot be verified in a real-life setting. The only available test data is the one obtained on splitting the available dataset into test and train datasets.” - ES6

“(...)because of the under-representation of positive cases in such data where positive events are rare, it is hard to construct an accurate classifier using ANFIS.” - ES9

“(...) increasing the size of the training set, taking more account of outliers and the parameter space viewed by the model, and potentially by exploring alternate inherently probabilistic classifiers.” - ES11

“A fundamental part of the machine learning models we are using is the use of a training set. By using this set, we are implicitly assuming that candidates are not overwhelmingly members of a scenario not represented in the FP training set, or some other rare case that is not expected or understood. The coverage of the training set for typical scenarios is discussed in Section 5.7.” - ES12

“Our model was built to recognize specific patterns in data and has only limited ability to recognize patterns beyond those it was trained to detect. This means that there may be unusual planetary systems out there that are interesting but would not be recognized by our model and would not be flagged as interesting or worthy of study” - ES16

“(...)our algorithm only detects transit-like signals, we do not introduce other possible false positives (e.g., eclipsing binaries) in our data set” - ES20

Identificar exoplanetas com certas características

Além dos exoplanetas distinguirem entre sim de tamanhos e composições, a sua relação com a estrela hospedeira também varia. A distância que esses planetas estão de suas estrelas e a duração de sua órbita são alguns exemplos. Além disso, sistemas planetários podem ter mais de um planeta, assim como o nosso sistema solar. Todas essas diferenças podem dificultar a detecção dos exoplanetas, trazendo assim, um desafio para os modelos inteligentes, assim como relatam os estudos ES12 e ES13.

Estrelas com mais de um planeta podem apresentar uma curva de luz mais longa e com variações dentro da própria curva. Isso acontece quando dois planetas transitam entre a estrela e o observador ao mesmo tempo tornando esse um desafio para as técnicas inteligentes. O ES13 relata que uma das suas limitações é justamente não conseguir identificar sistemas multiplanetários ou exoplanetas com longos períodos de trânsito. Os estudos ES24 e ES26 apresentam dificuldade para encontrar exoplanetas em estrelas menos brilhantes e exoplanetas muito pequenos comparado a sua estrela, respectivamente. Ambos relatam que o ruído no sinal é um grande limitador.

“No adjustment has been made for candidates in multiple systems, due to the complexity of the resulting probabilities and the difficulty of ascertaining how many TCEs on a given star are likely planets, and hence should be counted in any multiplicity effects.”- ES12

“We intend to address these issues as part of ongoing work in this project (...) this should also help with identifying both stars with multiplanet systems as well as planets with longer periods“ - ES13

“However, we do not expect candidates to increase exponentially with the number of targets, as transit events become more difficult to detect around dimmer targets, whose signals are relatively more contaminated with sources of noise.” - ES24

“The detection of such small Earth-sized planets are difficult because the transit depth, ~100 ppm for a solar type star, (...) By whitening the data we remove the scale of the transit depth from the observations such that the limiting detection factor is the noise parameter.” - ES26

Diferenciar Falsos positivos

Os estudos ES14, ES16, ES18 e ES20 mostram limitações em enfrentar os desafios de diferenciar falsos positivos de verdadeiros exoplanetas. Algumas das causas para esses falsos positivos são as diferentes características que as estrelas possuem. Existem estrelas chamadas de Estrelas variáveis, que são estrelas que variam sua luminosidade mesmo que não tenham nenhum objeto a bloqueando. Também existem binárias eclipsantes que são duas estrelas que orbitam entre si, causando variações na luz que chega no observador. Quando uma estrela sobrepõe a outra, o efeito de curva de luz acontece, o que pode ser confundido com um exoplaneta. Além disso, erros nos instrumentos de observação também podem causar falsos positivos.

“Kepler confuses eclipsing binary stars, non-eclipsing variable stars, instrumental artifacts or background events as coherent transit signal” - ES14

“(...) we might not be able to accurately simulate all types of instrumental and astrophysical false positives.” - ES16

“We observed that the linear and fully connected architectures have difficulty distinguishing U-shaped transits (like planets) from V-shaped transits (like eclipsing binaries)” - ES18

“our algorithm only detects transit-like signals, we do not introduce other possible false positives (e.g., eclipsing binaries) in our data set” - ES20

Qualidade dos dados

Outra limitação encontrada durante a análise destes estudos foi a qualidade dos dados utilizados. O estudo ES12 apresenta um desafio ao testar casos isolados de candidatos. Já que se os valores para esse caso forem ruins, os resultados obtidos futuramente vão ser enviesados. Já o estudo ES15 utilizou apenas dados de exoplanetas confirmados por seres humanos, originários da missão Kepler da NASA, e isso também pode enviesar o modelo.

“Issues with single candidates are however a potential problem, and an individual tested candidate with bad input values may lead to bad scores.” - ES12

“In the Kepler dataset, only those signals identified as planet candidates by humans were positive classes, introducing a possible human bias” - ES15

Na figura 16, é possível ver as técnicas de IA que foram utilizadas nos estudos para cada desafio ou limitação identificados.

Técnicas utilizadas	Cobertura do conjunto de dados do treinamento	Identificar exoplanetas com certas características	Diferenciar Falsos-Positivos	Qualidade dos dados
Convolutional Neural Network	4x	3x	3x	1x
Random Forest	3x	1x		1x
Multilayer perceptron	1x	2x		1x
Gaussian process classifier	1x	1x		1x
Extra trees	1x	1x		1x
Fully Connected Neural Network			1x	
Linear Neural Network			1x	
Support-vector machine	1x	1x		
AdaBoost	1x			
Feedforward neural network	1x			
Box fitting		1x		

method				
Wavelet MLP		1x		
Generative adversarial network	1x			
Adaptive neuro fuzzy inference system	1x			
Self-organizing map	1x			

Tabela 16: Técnicas utilizadas em cada desafio ou limitação identificados.

Fonte: Elaborado pelo autor

5.5.6 LIMITAÇÕES DA PESQUISA E AMEAÇA À VALIDADE

Para a execução deste trabalho foram utilizadas 6 bases de estudos diferentes (ACM, ADS, IEEE, ResearchGate, Science Direct e Scopus). Porém, utilizamos apenas os estudos que estavam disponíveis para acesso gratuito ou com acesso institucional, o que pode ter limitado o alcance da pesquisa, pois muitos estudos pagos e sem acesso institucional foram removidos da análise. Além disso, apesar das bases escolhidas serem de grande qualidade e importância no ramo da computação e da astronomia, existem outras bases de dados de qualidades, que poderiam agregar ainda mais valor e quantidade de estudos. Por fim, é importante ressaltar que mesmo que este trabalho considere estudos do ano de 2022, ele está limitado aos estudos do ano de 2022 publicados apenas nos primeiros 4 meses do ano.

6. CONCLUSÃO E TRABALHOS FUTUROS

6.1 CONCLUSÃO

Este estudo buscou trazer uma visão geral do atual cenário do uso da Inteligência artificial para a identificação de exoplanetas. Para isso, foi realizada uma revisão sistemática de literatura que foi capaz de reunir informações que mostraram quais as características e

contribuições das técnicas de inteligência artificial que contribuem para a identificação de exoplanetas. Para isso, os dados extraídos dos estudos responderam às seguintes perguntas: “Quais técnicas são utilizadas atualmente para ajudar na identificação de exoplanetas?”, “Quais os impactos e contribuições que essas técnicas trouxeram para a identificação de exoplanetas?”, “Quais técnicas apresentaram melhor desempenho?” e “Quais métodos de identificação de exoplanetas tiveram maior contribuição do uso de Inteligência artificial?”

A revisão sistemática de literatura conseguiu, por meio da estratégia de busca definida em 3.2, reunir 427 estudos, que passaram por avaliações de critérios de exclusão e inclusão e de qualidade. No final da etapa de seleção, 25 estudos foram selecionados para serem analisados.

Convolutional Neural Network (CNN), Random Forest, Support-vector machine e Multilayer perceptron foram as técnicas mais utilizadas nos estudos, foram utilizadas em 16, 8, 5 e 4 estudos respectivamente. Dentre as técnicas que obtiveram melhor desempenho em seus contextos específicos, a CNN foi a que se destacou. Tendo o melhor desempenho em 15 estudos. Random Forest obteve o melhor desempenho em 2 estudos.

Em relação aos métodos de identificação de exoplanetas que foram utilizados nos estudos, o método de trânsito foi o que se destacou. 88% dos estudos utilizaram técnicas de inteligência artificial em conjunto com este método. 2 estudos, ou seja, 8% utilizaram a inteligência artificial para detectar por meio de imagem direta e apenas 1 estudo utilizou o método de velocidade radial.

Após uma análise dos trabalhos selecionados, foi possível identificar que os estudos encontrados usaram a inteligência artificial em prol da detecção de exoplanetas de diversas formas, desde análise de dados de curvas de luz, até o uso de deep learning para análise de imagens 2D. Outra contribuição que foi feita em 9 dos estudos selecionados, foi a detecção de novos candidatos e validação de novos exoplanetas com a ajuda dos modelos de IA que foram implementados nos ES. Além disso, foram apresentados ganhos de eficiência, uso de modelos híbridos de IA e até foi possível analisar estrelas que antes eram pouco exploradas, até mesmo pelas missões espaciais da NASA.

Também analisamos criticamente os estudos com o objetivo de encontrar quais são as limitações e desafios que os estudos selecionados enfrentaram ao aplicar IA na identificação de exoplanetas. Foi identificado que em muitos estudos a cobertura do conjunto de dados do

treinamento e a qualidade dos dados utilizados eram fatores limitantes. E conseguir diferenciar falsos positivos, como binárias eclipsantes e estrelas variáveis, de verdadeiros exoplanetas e conseguir identificar exoplanetas de características que fogem do padrão, ainda são desafios ao utilizar modelos de IA para realizar esta atividade.

Portanto, conclui-se que este trabalho tem grande importância para a academia, pois fornece uma visão geral do cenário atual do uso da inteligência artificial na identificação de exoplanetas. Com esse trabalho, é possível saber quais as técnicas que estão sendo utilizadas e para qual finalidade, quais apresentaram melhor desempenho, para quais métodos de detecção de exoplanetas a IA está sendo utilizada, quais descobertas foram feitas, quais impactos e contribuições seu uso está causando e quais os limites e desafios que ainda podem ser superados futuramente pelo uso de IA na identificação de exoplanetas.

6.2 TRABALHOS FUTUROS

Com base no estudo realizado, sugerimos os seguintes direcionamentos de pesquisas para trabalhos futuros:

- Estender a pesquisa para um período de 10 anos;
- Estender a pesquisa para incluir o uso de técnicas de inteligência artificial para a busca por planetas em zona habitável;
- Estender a pesquisa para incluir mais bases de dados
- Estender a pesquisa para que sejam considerados estudos pagos
- Limitar a pesquisa ao uso de inteligência artificial apenas para a um determinado método de identificação de exoplanetas.
- Contribuir em projetos de pesquisa na área de detecção de exoplanetas, aprimorando técnicas de IA ou ainda alguma das etapas necessárias ao processo de decisão.

REFERÊNCIAS

[1] Budrikis, Z. 30 years of exoplanet detections. *Nat Rev Phys* (2022). <https://doi.org/10.1038/s42254-022-00459-x>.

[2] Wolszczan, A., Frail, D. A planetary system around the millisecond pulsar PSR1257 + 12. *Nature* 355, 145–147 (1992). <https://doi.org/10.1038/355145a0>.

[3] Cosmic Milestone: NASA Confirms 5,000 Exoplanets. *Exoplanet Exploration: Planets Beyond our Solar System*, 2022. Disponível em: <<https://exoplanets.nasa.gov/news/1702/cosmic-milestone-nasa-confirms-5000-exoplanets/>>. Acesso em: 28 de Apr. de 2022

[4] Pearson, K. A., Palafox, L., and Griffith, C. A. Searching for exoplanets using artificial intelligence, *Monthly Notices of the Royal Astronomical Society*, vol. 474, no. 1, pp. 478–491, 2018. doi:10.1093/mnras/stx2761.

[5] 5 Ways to Find a Planet. *Exoplanet Exploration: Planets Beyond our Solar System*, 2022. Disponível em: <<https://exoplanets.nasa.gov/alien-worlds/ways-to-find-a-planet/#>>. Acesso em: 30 de Apr. de 2022.

[6] EXOPLANET DETECTION METHODS. European Space Agency, 2019. Disponível em: <<https://sci.esa.int/web/exoplanets/-/60655-detection-methods>>. Acesso em: 30 de Abr. de 2022

[7] THE FUTURE OF EXOPLANET RESEARCH. European Space Agency, 2020. Disponível em: <<https://sci.esa.int/web/exoplanets/-/60657-the-future-of-exoplanet-research>>. Acesso em: 01 de Mai. de 2022

[8] Exoplanet Catalog | Discovery. *Exoplanet Exploration: Planets Beyond our Solar System*, 2020. Disponível em: <<https://exoplanets.nasa.gov/discovery/exoplanet-catalog/>>. Acesso em: 01 de Mai. de 2022

[9] What is an Exoplanet?. *Exoplanet Exploration: Planets Beyond our Solar System*, 2021. Disponível em: <<https://exoplanets.nasa.gov/what-is-an-exoplanet/overview/>>. Acesso em: 01 de Mai. de 2022

[10] Konacki, M., Torres, G., Jha, S. et al. An extrasolar planet that transits the disk of its parent star. *Nature* 421, 507–509 (2003). <https://doi.org/10.1038/nature01379>

[11] Haenlein M, Kaplan A. A Brief History of Artificial Intelligence: On the Past, Present, and Future of Artificial Intelligence. *California Management Review*. 2019;61(4):5-14. doi:10.1177/0008125619864925

[12] Dobrev D. A Definition of Artificial Intelligence. In: *Mathematica Balkanica, New Series*, Vol. 19, 2005, Fasc. 1-2, pp.67-74. <https://doi.org/10.48550/arXiv.1210.1568>

[13] BRERETON, Pearl et al. Lessons from applying the systematic literature review process within the software engineering domain. *Journal of Systems and Software*, Volume 80, Issue 4, April 2007, Pages 571-583.. Disponível em: <<https://doi.org/10.1016/j.jss.2006.07.009>>. Acesso em: 04 de Mar. de 2022.

[14] Xiao Y, Watson M. Guidance on Conducting a Systematic Literature Review. *Journal of Planning Education and Research*. 2019;39(1):93-112. doi:10.1177/0739456X17723971]. Acesso em: 04 de Mar. de 2022.

[15] Kitchenham, B. Charters, S. Guidelines for performing systematic literature reviews in software engineering. Citeseer, 2007. Disponível em: https://www.elsevier.com/_data/promis_misc/525444systematicreviewsguide.pdf. Keele University; Durham: University of Durham, 9 July 2007. Acesso em: 3 Mar. 2022.

[16] B. A. Kitchenham, E. Mendes and G. H. Travassos, "Cross versus Within-Company Cost Estimation Studies: A Systematic Review," in *IEEE Transactions on Software Engineering*, vol. 33, no. 5, pp. 316-329, May 2007, doi: 10.1109/TSE.2007.1001. acesso em : 05 de Mar. 2022.

[17] Jason T. Wright, B. Scott Gaudi. Exoplanet Detection Methods. Disponível em: <https://doi.org/10.48550/arXiv.1210.2471>. Planets, Stars and Stellar Systems. Volume 3: Solar and Stellar Planetary Systems, p. 489, 2012. Acesso em: 7 Mai. 2022.

[18] Robbins, Laura Pope. SAO/NASA Astrophysics Data System. Citeseer, 2012. Disponível em: <https://doi.org/10.5260/chara.20.2.42>. The Charleston Advisor, Volume 20, Number 2, 2018, pp. 42-46(5). Acesso em: 8 Mai. 2022.

[19] O'Shea, K. Nash, R. An Introduction to Convolutional Neural Networks. arXiv, 2015. Disponível em: <https://doi.org/10.48550/arXiv.1511.08458>. Acesso em: 18 Mai. 2022.

[20] Nascimento, E., Nguyen-Duc, A., Sundbø, I., and Conte, T., "Software engineering for artificial intelligence and machine learning software: A systematic literature review", 2020. Disponível em: <https://arxiv.org/pdf/2011.03751.pdf>. Acesso em 18 Mai. 2022.

[21] Watson, C., Cooper, N., Nader Palacio, D., Moran, K., and Poshyvanyk, D., "A Systematic Literature Review on the Use of Deep Learning in Software Engineering Research", 2020. Disponível em: <https://arxiv.org/pdf/2009.06520.pdf>. Acesso em 18 Mai. 2022.