



**UNIVERSIDADE ESTADUAL DA PARAÍBA
CAMPUS I - CAMPINA GRANDE
PRÓ-REITORIA DE PÓS-GRADUAÇÃO E PESQUISA
PROGRAMA DE PÓS-GRADUAÇÃO EM CIÊNCIA E TECNOLOGIA
AMBIENTAL
MESTRADO EM CIÊNCIA E TECNOLOGIA AMBIENTAL**

FELIPE BARROS DE SOUZA

**MODELO PREDITIVO DE INCÊNDIOS FLORESTAIS PARA O SEMIÁRIDO
NORDESTINO BASEADO EM REDES NEURAIS ARTIFICIAIS**

**CAMPINA GRANDE
2025**

FELIPE BARROS DE SOUZA

MODELO PREDITIVO DE INCÊNDIOS FLORESTAIS PARA O SEMIÁRIDO NORDESTINO BASEADO EM REDES NEURAIS ARTIFICIAIS

Dissertação apresentada à Coordenação do Curso de Mestrado Ciência e Tecnologia Ambiental da Universidade Estadual da Paraíba, como requisito parcial à obtenção do título de Mestre em Ciência e Tecnologia Ambiental.

Linha de Pesquisa: Qualidade de Sistemas Ambientais.

Orientador: Prof. Dr. William de Paiva

**CAMPINA GRANDE
2025**

É expressamente proibida a comercialização deste documento, tanto em versão impressa como eletrônica. Sua reprodução total ou parcial é permitida exclusivamente para fins acadêmicos e científicos, desde que, na reprodução, figure a identificação do autor, título, instituição e ano do trabalho.

S729m Souza, Felipe Barros de.

Modelo preditivo de incêndios florestais para o semiárido nordestino baseado em redes neurais artificiais [manuscrito] / Felipe Barros de Souza. - 2025.

76 f. : il. color.

Digitado.

Dissertação (Mestrado Ciência e Tecnologia Ambiental) - Universidade Estadual da Paraíba, Centro de Ciências e Tecnologia, 2025.

"Orientação : Prof. Dr. William de Paiva, Departamento de Engenharia Sanitária e Ambiental - CCT".

1. Incêndios Florestais. 2. Redes Neurais Artificiais.
3. Predição de incêndios. 4. Sensoriamento remoto. I.
Título

21. ed. CDD 628.92

FELIPE BARROS DE SOUZA

MODELO PREDITIVO DE INCÊNDIOS FLORESTAIS PARA O SEMIÁRIDO NORDESTINO BASEADO EM REDES NEURAIS ARTIFICIAIS

Dissertação apresentada à Coordenação do Curso de Mestrado Ciência e Tecnologia Ambiental da Universidade Estadual da Paraíba, como requisito parcial à obtenção do título de Mestre em Ciência e Tecnologia Ambiental

Linha de Pesquisa: Qualidade de Sistemas Ambientais.

Aprovada em: 28/08/2025.

BANCA EXAMINADORA

Documento assinado eletronicamente por:

- **Laércio Leal dos Santos** (***.451.154-**), em **27/09/2025 20:19:47** com chave **757896c29bf811f08b114ad2264da9bf**.
- **Stephanny Conceição Farias do Egito Costa** (***.897.684-**), em **29/09/2025 14:00:56** com chave **ddf1c5d09d5511f0a0c27ec8eff7bc99**.
- **ELISANGELA MARIA RODRIGUES ROCHA** (***.260.692-**), em **29/09/2025 09:02:58** com chave **3dc968e9d2c11f0b9407ec8eff7bc99**.
- **Marcia Ramos Luiz** (***.975.784-**), em **29/09/2025 08:19:07** com chave **1d6fe9889d2611f0b5e54ad2264da9bf**.
- **William de Paiva** (***.897.104-**), em **29/09/2025 12:55:47** com chave **c3b6fb949d4c11f0974ef2810dfffc275**.

Documento emitido pelo SUAP. Para comprovar sua autenticidade, faça a leitura do QrCode ao lado ou acesse https://suap.uepb.edu.br/comum/autenticar_documento/ e informe os dados a seguir.

Tipo de Documento: Folha de Aprovação do Projeto Final

Data da Emissão: 29/09/2025

Código de Autenticação: b5b24d



AGRADECIMENTOS

À minha Esposa Neyliane Costa de Souza, sua força, paciência e incentivo incansável foram o alicerce para que eu pudesse me dedicar aos estudos e ao aperfeiçoamento de meus conhecimentos. Esta conquista é tão minha quanto sua.

A minha Filha Carolina, que com seu carinho me acolhe, com sua curiosidade me inspira e com sua inocência me lembra do que realmente importa.

À minha Mãe Rosana Maria Barros de Souza, por ser grande responsável pela minha formação pessoal, sempre me incentivando a evoluir.

Ao meu orientador, Prof. Dr. William de Paiva, agradeço pelo apoio incondicional, pela orientação precisa e pela confiança depositada neste trabalho desde o início.

Aos membros da banca examinadora por aceitarem prontamente o convite para participar da defesa. Sou grato às valiosas contribuições da Profa. Dra. Marcia Ramos Luiz, da Profa. Dra. Elisângela Maria Rodrigues Rocha e da Profa. Dra. Stephanny Conceição Farias do Egito Costa.

Aos colegas de classe pelos momentos de amizade e aos colegas dos grupos de pesquisa GPTECA pela grande colaboração.

À Universidade Estadual da Paraíba (UEPB), pela excelência acadêmica e pela oportunidade de desenvolver esta pesquisa em seu Programa de Pós-Graduação em Ciência e Tecnologia Ambiental. Agradeço por toda a estrutura e suporte oferecidos.

Aos professores da Universidade Estadual da Paraíba, pela dedicação e disposição em cooperar na execução deste trabalho.

E a todos que contribuíram direta ou indiretamente no desenvolvimento deste trabalho.

“Porque foi tão fácil (não foi)
conseguir E agora eu me
pergunto: E daí?
Eu tenho uma porção
De coisas grandes pra
conquistar E eu não posso
ficar aí parado.”

(Raul Santos Seixas)

RESUMO

Os incêndios florestais representam uma grave ameaça ao equilíbrio ecológico e socioeconômico, especialmente no semiárido nordestino, uma região caracterizada por longos períodos de estiagem e vegetação suscetível ao fogo. A capacidade de prever a ocorrência de focos de incêndio é, portanto, uma ferramenta essencial para o planejamento de ações de mitigação e resposta rápida. Este trabalho teve como objetivo principal desenvolver um modelo preditivo para focos de incêndios florestais na região do Sertão de Crateús, Ceará, utilizando Redes Neurais Artificiais (RNA). Para isso, foram coletados dados de focos de calor de satélites (MODIS e VIIRS) e variáveis meteorológicas (temperatura, radiação solar, umidade, precipitação, vento e pressão atmosférica) do INMET, abrangendo o período de 2019 a 2024. A metodologia envolveu o processamento e a integração desses dados para treinar, validar e testar uma RNA do tipo Multi-Layer Perceptron (MLP). A análise de importância das variáveis, realizada por permutação de features, e a radiação solar global (RADGLB) foram os fatores de maior impacto na predição. O modelo final alcançou uma acurácia global de 72% nos dados de teste, com uma sensibilidade de 77%, indicando alta capacidade para identificar corretamente a ocorrência real de fogo. A análise das curvas de aprendizado demonstrou uma convergência estável com overfitting controlado. Os resultados indicam que a modelagem com RNA é uma abordagem promissora e eficaz para prever focos de incêndio, fornecendo uma base sólida para o desenvolvimento de sistemas de alerta precoce e auxiliando na gestão de recursos para o combate a incêndios no semiárido.

Palavras-chave: incêndios florestais; redes neurais artificiais; predição; semiárido; sensoriamento remoto.

ABSTRACT

Forest fires pose a serious threat to the ecological and socioeconomic balance, especially in the northeastern semiarid region of Brazil, an area characterized by long dry seasons and fire-susceptible vegetation. The ability to predict the occurrence of fire hotspots is, therefore, an essential tool for planning mitigation actions and rapid response. This study aimed to develop a predictive model for forest fire hotspots in the Sertão de Crateús region, Ceará, using Artificial Neural Networks (ANN). For this purpose, data on fire hotspots from satellites (MODIS and VIIRS) and meteorological variables (temperature, solar radiation, humidity, precipitation, wind, and atmospheric pressure) from INMET were collected, covering the period from 2019 to 2024. The methodology involved processing and integrating these data to train, validate, and test a Multi-Layer Perceptron (MLP) type ANN. The analysis of variable importance, carried out using feature permutation, revealed that maximum atmospheric pressure (PATMAX) and global solar radiation (RADGLB) were the factors with the greatest impact on the prediction. The final model achieved an overall accuracy of 72% on the test data, with a sensitivity of 77%, indicating a high capacity to correctly identify actual fire occurrences. The analysis of the learning curves demonstrated stable convergence with controlled overfitting. The results indicate that ANN modeling is a promising and effective approach for predicting fire hotspots, providing a solid basis for the development of early warning systems and assisting in resource management for firefighting in the semiarid region.

Keywords: forest fires; artificial neural networks; prediction; semiarid; remote sensing.

LISTA DE ILUSTRAÇÕES

Figura 1	Tríade do fogo e Tetraedro do fogo	18
Figura 2	Esquema comparativo entre (A) neurônio biológico e (B) arquitetura de rede neural artificial	28
Figura 3	Arquitetura de Perceptron de Camada Simples	29
Figura 4	Arquitetura de Perceptron de Múltiplas Camadas (MLP)	30
Figura 5	Fluxograma das etapas metodológicas do estudo	34
Figura 6	Região do Estudo e abrangência na região do sertão de Crateús, CE	36
Figura 7	Área de Estudo com destaque para a bounding box, Microrregião do Sertão de Crateús, Ceará	36
Figura 8	Fases do processamento de dados e as etapas de validação e caracterização	41
Figura 9	Representação esquemática da RNA com uma camada oculta	45
Figura 10	Gráfico das variáveis influentes para previsão de ocorrência de incêndios florestais na região de estudo	50
Figura 11	Matriz de confusão obtida durante o treinamento da rede neural	56
Figura 12	Gráficos de Acurácia e Função de Perda durante o Treinamento	57
Figura 13	Comparação de desempenho entre o Modelo Completo e o Modelo Reduzido	60

LISTA DE TABELAS

Tabela 1	Estimativa da População, Área, Índice Pluviométrico, coordenadas e clima da região em estudo	37
Tabela 2	Relação sintetizada de dados de médias mensais de variáveis meteorológicas para o segundo semestre de 2024	38
Tabela 3	Comparação entre os modelos de RNA.	55
Tabela 4	Relatório de Classificação	58
Tabela 5	Comparação de métricas de desempenho entre o modelo completo e o modelo reduzido	60

LISTA DE ABREVIATURAS E SIGLAS

- ABNT: Associação Brasileira de Normas Técnicas
ANA: Agência Nacional de Águas
ANN: *Artificial Neural Networks* (Redes Neurais Artificiais)
ARIMA: Modelos Autorregressivos de Médias Móveis Integradas
CCT: Centro de Ciências e Tecnologia
CPTEC: Centro de Previsão de Tempo e Estudos Climáticos
DIRVENT: Direção do Vento
FAO: *Food and Agriculture Organization of the United Nations*
FIRMS: Fire Information for Resource Management System
FRP: Poder Radioativo do Fogo
INMET: Instituto Nacional de Meteorologia
INPE: Instituto Nacional de Pesquisas Espaciais
ISO: *International Organization for Standardization*
JPSS: Joint Polar Satellite System.
MLP: *Multi-Layer Perceptron* (Perceptron de Múltiplas Camadas)
MODIS: *Moderate Resolution Imaging Spectroradiometer*
PAH: Hidrocarbonetos Aromáticos Policíclicos
PATMAX: Pressão Atmosférica Máxima
PATMIN: Pressão Atmosférica Mínima
PATMNE: Pressão Atmosférica ao Nível da Estação
PM: Material Particulado
PPGCTA: Programa de Pós-Graduação em Ciência e Tecnologia Ambiental
PRECIP: Precipitação
RADGLB: Radiação Solar Global
RAJMAX: Rajada Máxima do Vento
RL: Regressões Lineares
RNA: Redes Neurais Artificiais
SGD: Stochastic Gradient Descent
SIG: Sistemas de Informação Geográfica

TEMPBUL: Temperatura do Ar Bulbo Seco

TEMPMAX: Temperatura Máxima

TEMPPMIN: Temperatura Mínima

TEMPORV: Temperatura do Ponto de Orvalho

TMAXORV: Temperatura Máxima do Ponto de Orvalho

TMINORV: Temperatura Mínima do Ponto de Orvalho

UEPB: Universidade Estadual da Paraíba

UESC: Universidade Estadual de Santa Cruz

UFPB: Universidade Federal da Paraíba

UMIDREL: Umidade Relativa

UMIRMAX: Umidade Relativa Máxima

VELVENT: Velocidade do Vento

VIIRS: *Visible Infrared Imaging Radiometer Suite*

VLSI: *Very-Large-Scale-Integration* (Integração em Larga Escala)

LISTA DE QUADROS

Quadro 1	Descrição dos componentes do fogo	17
Quadro 2	Pontos de destaque nas pesquisas da área de estudo	25
Quadro 3	Resumo da Metodologia	42

SUMÁRIO

1	INTRODUÇÃO	14
2	OBJETIVOS.....	16
3	FUNDAMENTAÇÃO TEÓRICA	17
3.1	Incêndio Florestal.....	17
3.1.1	Agentes Causadores de Incêndios.....	19
3.1.2	Incêndios no Bioma Caatinga.....	21
3.1.3	Exposição de Poluentes Oriundos de Incêndios	22
3.2	Predição e Monitoramento de Incêndios.....	23
3.2.1	Estudos Aplicados na Predição de Incêndios.....	24
3.3	Redes Neurais Artificiais	26
3.3.1	Definição e Fundamentos Teóricos	26
3.3.2	Arquiteturas Fundamentais de RNAs	28
3.3.3	Mecanismo de Aprendizado e Propriedades Fundamentais	30
3.3.4	Paradigmas de Aprendizado e Algoritmos	31
3.3.5	Funções de Ativação	32
4	MATERIAIS E MÉTODOS	34
4.1	Área de Estudo	35
4.2	Coleta de Dados	37
4.2.1	Fontes de Dados Meteorológicos	37
4.2.2	Processamento de Dados Climatológicos	43
4.2.3	Preparação e Análise de Dados para Modelagem	43
4.2.4	Arquitetura da RNA	44
4.2.5	Treinamento e Configuração do Modelo	45
5	RESULTADOS E DISCUSSÃO	47
5.1	Análise Inicial com Todas as Variáveis	47
5.2	Análise da Importância das Variáveis	49
5.2.1	Radiação Global	50
5.2.2	Temperatura Máxima	51
5.2.3	Velocidade do Vento	52
5.2.4	Umidade Relativa	52
5.2.5	Precipitação.....	53
5.3	Estratégia de Simplificação do Modelo	53

5.3.1 Resultados do Modelo Simplificado	54
5.3.2 Matriz de Confusão	55
5.3.3 Curva de Aprendizado.....	56
5.3.4 Relatório de Classificação Detalhado.....	58
5.4 Comparação de Desempenho: Modelo Completo vs. Modelo Reduzido	59
5.5 Implicações Práticas e Usabilidade do Modelo	61
5.5.1 Facilidade de Coleta e Disponibilidade de Dados	61
5.5.2 Interpretabilidade e Tomada de Decisão.....	61
5.5.3 Eficiência Operacional e Escalabilidade.....	62
5.5.4 Potencial para Integração em Sistemas de Alerta Precoce.....	62
5.5.5 Contribuição para a Gestão Ambiental e Políticas Públicas	62
6 CONSIDERAÇÕES FINAIS	64
REFERÊNCIAS.....	66
APÊNDICE A - CÓDIGO DO ALGORITMO DA REDE NEURAL FINAL.....	72

1 INTRODUÇÃO

As florestas desempenham um papel crucial na preservação do equilíbrio ecológico do planeta. Fatores naturais e humanos podem causar incêndios, e estes incêndios representam uma ameaça significativa aos valiosos recursos naturais.

Os incêndios florestais são definidos como toda e qualquer energia derivada da combinação entre combustível, comburente e calor, capaz de consumir vegetações vivas ou mortas. Esses incêndios, podem iniciar-se de forma espontânea ou ser consequência de ações humanas, por conseguinte, é iminente a probabilidade de ocorrências de incêndios florestais em todo território brasileiro, principalmente na queima de biomassa vegetal.

A intensificação da queima de biomassa, comum em diversas culturas agrícolas, acarreta graves prejuízos ecológicos, econômicos e paisagísticos. Essa prática afeta diretamente os biomas brasileiros, cuja maioria apresenta algum grau de vulnerabilidade ao fogo. Entre eles, o bioma Caatinga se destaca como uma das áreas mais suscetíveis a grandes incêndios (Jesus *et al.*, 2020).

Além da destruição em massa de diferentes ecossistemas, das alterações climáticas, das perdas ambientais, econômicas e sociais, os incêndios florestais também são responsáveis pela emissão de quantidades significativas de gases poluentes (Fiedler *et al.*, 2023). A queima de biomassa constitui uma das maiores fontes de aerossóis carbonáceos (carbono orgânico; carbono elementar; carbono negro) e gases residuais na atmosfera global (Vadrevu *et al.*, 2015; de Sa *et al.*, 2019; Kalita *et al.*, 2020; Heinold *et al.*, 2022), contribuindo com cerca de 42% e 74% das emissões de aerossóis carbonáceos e gases residuais, respectivamente (Bond *et al.*, 2013; Zhu *et al.*, 2022). Ademais, os aerossóis provenientes dos incêndios florestais podem impactar negativamente o clima regional através da atenuação da radiação solar e da modificação das propriedades atmosféricas (Jiang *et al.*, 2020; Yang *et al.*, 2022; Dumka *et al.*, 2022; Zhu *et al.*, 2022; Zhu *et al.*, 2022).

No Brasil, foram determinados indicadores da qualidade do ar, representados pelos principais gases poluentes, sendo eles: partículas inaláveis, dióxido de enxofre, dióxido de nitrogênio, monóxido de carbono e ozônio

(INSTITUTO ESTADUAL DO AMBIENTE; GERÊNCIA DE QUALIDADE DO AR, 2016).

A execução de previsões sobre focos de queimadas e incêndios, é uma atividade complexa, que depende de um conjunto de informações de satélites, análise de dados históricos, de estudos sobre a vegetação, composição do solo e de diversas outras variáveis ambientais. Outrossim, dados precisos sobre as variáveis ambientais, podem ser obtidos através do sensoriamento remoto, aliado aos sistemas de informações geográficas, às técnicas de inteligência artificial e estatística aplicada, favorecendo às tomadas de decisão na previsão.

Nessa conjuntura, a previsão de focos de incêndios utilizando inteligência artificial amplia significativamente a obtenção de dados em locais que periodicamente apresentam comprometimento dos biomas e piora da qualidade do ar.

Neste contexto, com a crescente frequência e gravidade dos incêndios florestais em todo o mundo, práticas eficazes de gestão de incêndios que integrem tecnologias inteligentes são cruciais para mitigar os impactos dos incêndios florestais.

Este trabalho propõe a predição de focos de incêndios em áreas suscetíveis a queimadas, através da aplicação da metodologia de Redes Neurais Artificiais(RNAs).

2 OBJETIVOS

O presente trabalho tem como objetivo utilizar RNAs para prever focos de incêndio, visando auxiliar tanto na identificação de áreas de risco quanto na aplicação de possíveis medidas preventivas.

Objetivos Específicos

Para a consolidação do objetivo geral, faz-se necessário realizar os seguintes objetivos específicos, como etapas pertinentes ao estudo:

- Identificar e analisar as variáveis climáticas que contribuem para o aumento do risco de incêndios no cenário em estudo;
- Desenvolver e adaptar uma rede neural artificial (RNA) e validar o modelo;
- Avaliar o desempenho do modelo de RNA para classificar a ocorrência de focos de incêndio, validando sua eficácia preditiva.

3 FUNDAMENTAÇÃO TEÓRICA

3.1 Incêndio Florestal

Incêndio é uma ocorrência de fogo não controlado, e pode ser extremamente perigoso para os seres vivos e estruturas. Considera-se como incêndios florestais a ocorrência do fogo em vegetação, sem controle e com potencial destrutivo, cuja incidência advém de causas naturais ou antrópicas, sendo estas últimas apontadas em muitas pesquisas como a principal origem dessas queimadas (Ramalho *et al.*, 2021; Silva, 2017).

Durante muitos anos o triângulo do fogo (combustível, comburente e calor ou energia térmica) foi utilizado para ensinar os componentes do fogo (Quadro 1). Porém, se não houver condições ideais, ou seja, a presença simultânea e proporcional dos três componentes, não haverá sustentabilidade do fogo. Por isso, foi acrescentado ao triângulo do fogo (Figura 1) uma quarta face, denominada reação em cadeia, que interliga todos os elementos promovendo a existência e a continuidade do fogo (Corpo de Bombeiro Militar de Santa Catarina, 2018).

Quadro 1 - Descrição dos componentes do fogo.

Componente	Denominação
Combustão	Ação exotérmica de uma substância combustível com um oxidante, usualmente acompanhada por chamas e/ou abrasamento e/ou emissão de fumaça.
Fogo	processo de combustão caracterizado pela emissão de calor acompanhado por fumaça, chama ou ambos.
Combustível	Fornece energia para a queima, representado por elementos susceptíveis a entrar em combustão como: madeira, papel, pano, estopa, entre outros.
Comburente	Substância que reage quimicamente com o combustível e inicia a combustão, sendo o principal comburente o oxigênio.
Fonte de calor	Necessário para iniciar a reação, pode ser desde luz solar, faíscas de raios atmosféricos a pontas de cigarro.

Fonte: International Organization for Standardization (ISO)

Figura 1 – Tríade do fogo e Tetraedro do fogo



Fonte: gestaodesegurancaprivada.com.br (2021)

Os incêndios desempenham um papel importante na regulação do clima regional e global, perturbando os processos meteorológicos, biogeoquímicos e hidrológicos (Zou *et al.*, 2019). Outro agravante está associado às questões culturais de uso da terra, como por exemplo, a prática de queimadas para manejo da pastagem e preparo do solo para plantio de culturas agrícolas. O fogo também é usado para manejo da pecuária, em savanas e campos naturais, para remoção do excesso de biomassa e estímulo da rebrota (Pivello *et al.*, 2021). De acordo com o relatório descrito pela Food and Agriculture Organization of the United Nations - FAO (2006), sobre incêndios na América do Sul em até 90% dos incêndios florestais são causados por atividades humanas.

No Brasil, conforme Oliveira Júnior *et al.* (2017), os números de focos desses incêndios e de queimadas aumentaram significativamente nos últimos vinte anos, ocasionando considerável interesse por eventos que dizem respeito ao tema. Nessa perspectiva, conforme os dados do Instituto Nacional de Pesquisas Espaciais (INPE), no ano de 2019 houve crescimento de 48% no total de focos de incêndios no Brasil em relação ao ano anterior, enquanto em 2020 ocorreu elevação de 12% em comparação com 2019.

Um relatório recente do Banco Mundial identificou que o Brasil está entre os quatro países do mundo que mais utilizam o fogo para a queima de resíduos agrícolas. Estudos também indicam que existe uma forte relação entre anos de secas extremas e impactos na saúde devido à ocorrência de queimadas e incêndios florestais, como em anos de El Niño (Cassou, 2018).

Um outro fator importante sobre Incêndios Florestais, versa sobre os impactos econômicos, um estudo realizado por Silva (2022), apontou os recursos gastos pelas instituições: Corpo de Bombeiros, Exército, Núcleo de Operações e Transporte Aéreo e o ICMbio, resultando num gasto operacional total de R\$ 358.438,97, com recursos empregados em 21 dias de combate aos incêndios.

Os incêndios naturais e antropogênicos desempenham um papel fundamental no ciclo do carbono terrestre, e são uma importante fonte de emissões na atmosfera de gases do efeito de estufa, como o monóxido de carbono, aerossóis carbonáceos e uma série de outros gases e partículas (Adame *et al.*, 2018; Akagi *et al.*, 2011; Miranda *et al.*, 2008; Van Der Werf *et al.*, 2010).

3.1.1 Agentes Causadores de Incêndios

Segundo a Organização das Nações Unidas para Agricultura e a Alimentação - FAO (2006), as principais causas dos incêndios florestais são:

- Raios: causados diretamente por diversas descargas elétricas da atmosfera. Não existe responsabilidade humana.
- Incendiários: provocados potencialmente por pessoas em propriedades alheias, seja por vingança ou por desequilíbrio mental.
- Queima para limpeza: originados do uso do fogo na limpeza do terreno para fins agrícolas, florestais ou pecuários, que por negligência ou descuido escapam do controle e atingem áreas florestais.
- Fumantes: provocados por fósforo ou por cigarros acesos.
- Operações florestais: provocados por trabalhadores florestais em atividades na floresta.
- Fogos de recreação: incêndios causados por pessoas que utilizam a floresta como local de recreação.
- Estradas de ferro: incêndios causados diretamente ou indiretamente pelas atividades de ferrovias.
- Diversos: incêndios com causa pouco frequente ou regionais não são enquadrados em uma classificação especial. Ex.: queda de aviões, incêndio de automóveis ou balões em festas juninas.

Parques *et al.*, (2018) investigaram as principais causas de incêndios de alta gravidade em ecorregiões florestais da região oeste dos Estados Unidos entre 2002 e 2015, descobrindo que o combustível vivo foi o principal fator (53,1%) na causa desses incêndios, seguido pelo clima de incêndio (22,9%). A topografia (10,3%) e o clima (13,7%) tiveram impacto menor.

Os fatores associados às mudanças climáticas, como a diminuição da precipitação, o aumento das temperaturas e os períodos prolongados de seca, juntamente com os efeitos prejudiciais das atividades humanas, exacerbaram o risco de incêndios florestais em várias áreas (Tien Bui *et al.*, 2018).

Bhatt *et al.*, (2023) citam como principais elementos influenciadores de incêndios florestais os fatores topográficos como elevação, aspecto e declive, variáveis climáticas como precipitação anual, velocidade do vento, índice de seca, temperatura e evapotranspiração, e um por último antropogênicos, nomeadamente a distância às estradas. Os incêndios florestais, especialmente os de grandes proporções, são o produto de dois fatores-chave que interagem entre si: suprimento de combustível e padrões climáticos. Nos últimos anos, foram feitos esforços significativos para estudar os regimes de incêndio históricos e atuais e avaliar a importância dos principais fatores nesses regimes.

É notável a existência de uma relação potencial entre variáveis meteorológicas e incêndios florestais (Tošiý *et al.*, 2019) aplicaram fatores meteorológicos para construir um índice de incêndios para avaliar a gravidade dos incêndios florestais na Sérvia. Já, Masinda *et al.*, (2022) utilizaram os indicadores meteorológicos internacionais para avaliar a natureza selvagem das florestas no Nordeste da China.

Estudos realizados (Butler *et al.*, 2020; Moon *et al.*, 2019), demonstraram que a velocidade do vento numa direção favorável contribuirá para a propagação do incêndio. Nelson *et al.*, (2002) estudaram os efeitos de diferentes velocidades do vento em modelos de propagação florestal, e a temperatura da superfície terrestre, a precipitação e a velocidade média do vento foram identificadas como indicadores secundários para análise posterior.

Adicionalmente, em um relatório oficial da Sérvia mostrou que as alterações na temperatura da superfície terrestre e na precipitação foram fatores importantes no comportamento do fogo florestal (Živanoviý *et al.*, 2020).

Por essa razão, o desafio tanto da prevenção quanto da supressão é antecipar e reduzir o potencial de propagação de grandes incêndios florestais, assim como o risco subsequente para vidas humanas, propriedades e sistemas de uso do solo (Tyndall, 2023).

3.1.2 Incêndios no Bioma Caatinga

O bioma Caatinga compõe uma região de rica biodiversidade, que a caracteriza como o semiárido mais biodiverso do mundo (SEYFFARTH; RODRIGUES, 2017). Esse bioma, tem seu domínio estendido por cerca de 912.529 km², o que corresponde aproximadamente a 11% do território nacional, presente em 10 estados brasileiros: Alagoas, Bahia, Ceará, Maranhão, Minas Gerais, Paraíba, Pernambuco, Piauí, Rio Grande do Norte e Sergipe.

O bioma Caatinga é caracterizado por uma floresta sazonal seca com uma longa estação de estiagem e temperaturas médias elevadas, condições que o tornam naturalmente suscetível ao fogo (SOUZA *et al.*, 2017). Embora a vulnerabilidade natural exista, são as atividades antrópicas as principais responsáveis pelos focos de incêndio, com registros associados a queimas para limpeza de pastagens, eliminação de restos vegetais e outras práticas que influem diretamente na propagação do fogo (Alvez *et al.*, 2021).

Essa pressão humana tem alterado o regime histórico do bioma. Conforme indicam estudos de longo prazo de Alencar *et al.* (2022), a Caatinga tem registrado uma concentração de incêndios na estação seca, entre julho e outubro. Portanto, a combinação das características fisiológicas e biogeográficas com as fontes de ignição humanas torna a Caatinga um ambiente propenso a incêndios, o que justifica a urgência no desenvolvimento de modelos preditivos, como o proposto neste trabalho, para o monitoramento do risco.

Silva (2021), observou que em todo o bioma Caatinga, os estados que apresentaram estatisticamente o maior número médio de ocorrência de fogo para os anos estudados foram o Piauí (3592,5), a Bahia (1491,89) e o estado do Ceará (838,56). Para os demais estados não foram observadas diferenças estatísticas entre si. As maiores áreas médias queimadas foram registradas no estado do Piauí (65,20 ha), seguido do estado da Bahia (55,71 ha) e Minas Gerais (54,11 ha). Em contrapartida, o estado de Sergipe (31,05 ha) e Alagoas (31,63ha) apresentaram as menores médias de área queimada.

3.1.3 Exposição de Poluentes Oriundos de Incêndios

Os impactos de poluentes atmosféricos na saúde já são conhecidos, e afetam principalmente as populações vulneráveis, como crianças, mulheres grávidas e idosos, e tem sido associada à mortalidade por todas as causas, principalmente por problemas cardíacos e respiratórios (Chen e Hoek, 2020; Hvidtfeldt. *et al.*, 2019).

Vários estudos têm relatado o impacto na saúde da exposição de curta duração, a partículas em suspensão com um diâmetro inferior a 2,5 micrômetros (PM 2,5) emitidas pelos incêndios, influenciando na mortalidade, em prejuízos cardiorrespiratórios e em internamentos hospitalares.

Wettstein *et al.* (2018) descreveram uma associação entre visitas aos hospitais devido a doenças cardíacas, cerebrovasculares e respiratórias e a densidade de fumaça de incêndios florestais na Califórnia, enquanto, Arriagada *et al.* (2020) estimaram o número de hospitalizações relacionadas a doenças cardíacas e respiratórias, mortes e atendimento hospitalar relacionadas à asma, relacionadas à exposição de curto prazo ao PM2,5, devido à fumaça de incêndios florestais na Austrália.

Jones *et al.* (2020) analisaram a associação entre exposição ao PM devido a incêndios florestais e parada cardíaca extra-hospitalar na Califórnia (EUA). Stowell *et al.* (2019) estimaram a ligação entre a exposição de curto prazo ao PM2,5 e eventos cardiorrespiratórios agudos no Colorado, usando um modelo de exposição que relaciona visitas ao departamento de emergência e hospitalizações por casos de doenças cardio respiratória aguda.

Ravi *et al.* (2019) usaram o Programa de Mapeamento e Análise de Benefícios Ambientais para calcular a mortalidade adicional por todas as causas causada pela exposição a PM2,5 em incêndios prescritos na região noroeste do Pacífico dos Estados Unidos, estimando que apenas os incêndios prescritos causaram 280 a 700 mortes adicionais.

Huang *et al.* (2019) quantificaram o impacto dos fogos prescritos na saúde humana, na Geórgia (EUA) durante as épocas de incêndios de 2015-2018, através dos atendimentos de urgência relacionadas com a asma, estimando um aumento no número de urgências por asma devido aos impactos dos incêndios.

Shi *et al.* (2019) avaliaram o impacto na qualidade do ar dos incêndios florestais ocorridos em dezembro de 2017 no sul da Califórnia, utilizando o modelo Weather Research and Forecasting with Chemistry, combinado com observações de satélite e de superfície. Os resultados mostraram que as concentrações de PM_{2,5} aumentaram significativamente, ultrapassando os limites de qualidade do ar dos EUA, indicando que este incêndio contribuiu para a exposição aguda e cumulativa de PM_{2,5} nesta região, o que pode causar morte prematura e efeitos cardiovasculares na população exposta (Shi *et al.*, 2019).

Zhang *et al.* (2023) também avaliaram o efeito da exposição a PM_{2,5} relacionada a incêndios florestais, mas em mulheres grávidas, indicando um aumento no número de nascimentos prematuros ou baixo peso ao nascer. Em Portugal (Esteves *et al.*, 2021; Miranda *et al.*, 2012; Oliveira *et al.*, 2016, 2020). Oliveira e cols. (2020) estimaram o impacto da exposição de curto prazo ao PM₁₀ nos sintomas asmáticos em crianças asmáticas, e ao PM_{2,5} no número de internações hospitalares por doenças cardiovasculares, bem como na mortalidade em adultos (todas causas naturais), mostrando um aumento durante esses episódios.

Os padrões de qualidade do ar adotados na legislação brasileira, são estabelecidos na resolução nº 03 do Conselho Nacional de Meio Ambiente – Conama (1990), que estabelece limites aceitáveis apenas para o PM₁₀ que é de 60 µm.m⁻³ para a média diária. Embora seja de fundamental importância para a avaliação de impactos da poluição atmosférica por queimadas na saúde humana, poucos estudos têm investigado diretamente os efeitos do PM_{2,5} na saúde humana.

3.2 Predição e Monitoramento de Incêndios

O Instituto Nacional de Pesquisas Espaciais (INPE) tem, desde a década de 80, aprimorado seu sistema de detecção de queimadas no Brasil, com foco na Amazônia a partir de 1998. Essa atividade se alinha aos objetivos do instituto de desenvolver tecnologias e produtos de utilidade social, como monitorar focos de incêndio via satélite, prever riscos de queima da vegetação e estimar emissões. Segundo o próprio INPE (2023), o monitoramento por imagens de satélite é, desse modo, particularmente útil para regiões remotas sem outros

meios de acompanhamento, situação comum em grande parte do território nacional.

Os focos de queimadas passaram a ser obtidos por imagens dos satélites NOAA (National Oceanic and Atmospheric Administration) quatro vezes ao dia e mais recentemente também nas imagens do GOES-Leste (Geostationary Operational Environmental Satellite) oito ou mais vezes ao dia, e TERRA e AQUA duas vezes por dia cada, sendo em seguida integrados a dois sistemas de informações geográficas acessíveis na Internet.

Em países como a Austrália, as informações sobre saúde pública são relacionadas com incêndios florestais e está disponível para o público em geral através do Ministério da Saúde daquele país, particularmente, sobre aqueles Estados que possuem histórico de episódios de queimadas.

Diferentes modelos têm sido comumente aplicados na predição desses eventos, incêndios florestais e queimadas, com a utilização de variáveis meteorológicas e de séries históricas de dados sobre queimadas e incêndios florestais, que podem ser descritos por séries temporais, modelos de regressões lineares (RL), modelos autorregressivos de médias móveis integradas (ARIMA) e as RNAs.

Portanto, o estudo de séries temporais tem aplicações em diversas áreas do conhecimento, uma vez que permite entender a natureza estocástica de fenômenos, identificar padrões e prever valores futuros a partir de um histórico conhecido (Morettin e Toloi, 1981).

3.2.1 Estudos Aplicados na Predição de Incêndios

Li et al. (2016) conduziram uma pesquisa notável que se concentrou no uso de dados de sensoriamento remoto para detecção de incêndio. O estudo avaliou vários algoritmos, como máquinas de vetores de suporte, RNAs e árvores de decisão, e relatou que esses métodos possuem um potencial significativo para melhorar a precisão da detecção de incêndio.

Chang et al (2012) dirigiram uma das primeiras revisões abrangentes sobre o uso de algoritmos no gerenciamento de incêndios, cobrindo tópicos como modelagem de comportamento de incêndio, previsão de propagação de incêndio e sistemas de apoio à decisão de gerenciamento de incêndio. O estudo sugere que a IA tem o potencial de melhorar as práticas de gestão de incêndios,

oferecendo previsões de incêndios mais precisas e confiáveis e ferramentas de apoio à decisão.

Teixeira *et al.* (2020) administraram um estudo recente que enfoca a aplicação de algoritmos no contexto de preparação e resposta a incêndios florestais. O estudo fornece uma revisão completa de vários algoritmos, como florestas aleatórias, árvores de decisão e RNAs, e destaca seu potencial para aumentar a eficácia dos esforços de preparação e resposta durante um incêndio florestal. Estas pesquisas demonstram o interesse crescente na aplicação destes no campo da ciência e gestão de incêndios florestais e destacam o potencial destes métodos para melhorar as práticas de gestão de incêndios.

As estratégias de prevenção de incêndios florestais para detecção e supressão melhoraram significativamente ao longo dos anos, tanto devido às inovações tecnológicas quanto à adoção de diversas competências e métodos. Hoje em dia, os investigadores de incêndios florestais utilizam tecnologias que integram dados sobre previsão meteorológica, topografia, modelagem de combustível e outros fatores para prever como os incêndios se propagam (Zacharakis e Tsirhrintzis, 2023; Bakhshaii e Johnson, 2019). O Quadro 2 apresenta as pesquisas em destaque na área de estudo.

Quadro 2 – Pontos de destaque nas pesquisas da área de estudo

Autor (Ano)	Foco Principal do Estudo
Viganó <i>et al.</i> (2017)	Previsão de focos de incêndio no Pantanal com Redes Neurais Artificiais (RNA-MLP).
Ghorbanzadeh <i>et al.</i> (2019)	Mapeamento de suscetibilidade a incêndios utilizando RNA-MLP e múltiplas variáveis ambientais.
Luz <i>et al.</i> (2023)	Análise experimental do comportamento do fogo em espécies da Caatinga.
Barboza <i>et al.</i> (2021)	Previsão de incêndios no Cerrado com dados meteorológicos e aprendizado de máquina.
Parente <i>et al.</i> (2021)	Análise dos padrões e dos impulsionadores de áreas queimadas na Caatinga.

Jain <i>et al.</i> (2020)	Revisão abrangente sobre o uso de aprendizado de máquina na ciência de incêndios.
Rodrigues <i>et al.</i> (2024)	Avaliação da suscetibilidade a incêndios em escala global com aprendizado de máquina.
TOŠIĆ <i>et al.</i> (2019)	Relação entre variáveis meteorológicas e incêndios florestais.
ŽIVANOVIĆ <i>et al.</i> (2020)	Influência da temperatura e precipitação no comportamento do fogo.
BHATT <i>et al.</i> (2024)	Estudo dos principais elementos que influenciam incêndios.
LI <i>et al.</i> (2016)	Uso de sensoriamento remoto para detecção de incêndios.
TEIXEIRA <i>et al.</i> (2020)	Aplicação de algoritmos para preparação e resposta a incêndios.
ZACHARAKIS et al. (2020)	Previsão de incêndios com dados de meteorologia e topografia.
CHANG <i>et al.</i> (2012)	Revisão sobre o uso de algoritmos em gerenciamento de incêndios.

Fonte: Elaborado pelo autor, 2025.

3.3 Redes Neurais Artificiais

3.3.1 Definição e Fundamentos Teóricos

A inteligência artificial (IA) foi desenvolvida a partir da necessidade de se analisar um grande volume de dados através da computação de alto desempenho, onde a ferramenta neural é uma das direções da inteligência artificial.

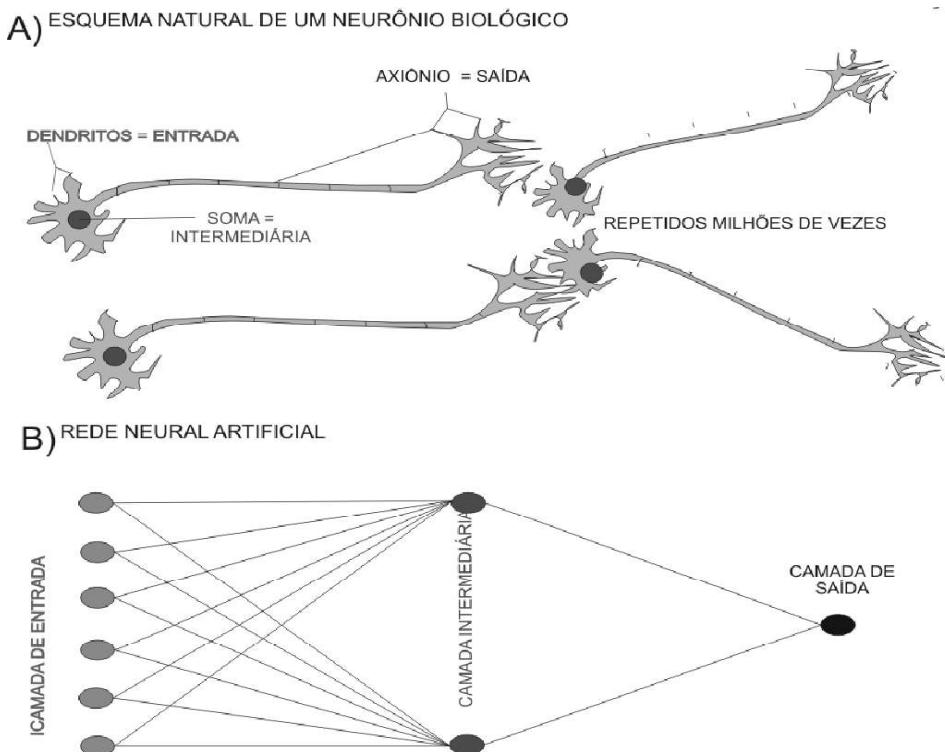
As RNAs são modelos computacionais, inspirados no sistema nervoso de seres vivos. A primeira rede neural foi concebida por Warren McCulloch e Walter Pitts em 1943. Eles escreveram um artigo seminal constituindo uma analogia entre as células nervosas e o processo eletrônico, em um artigo publicado no Bulletin of Mathematical Biophysics com o título: A Logical Calculus of the Ideas Immanent in Nervous Activity (Cálculo lógico de Ideias inerentes a Atividade nervosa). Assim, modelaram suas concepções criando uma rede neural simples com circuitos elétricos.

Numa relação entre neurônios e redes neurais, temos que os dendritos foram substituídos por entradas, cujas ligações com o corpo celular artificial são realizadas através de elementos chamados de peso que simulam as sinapses. Os estímulos captados pelas entradas são processados pela função de soma, e o limiar de disparo do neurônio biológico foi substituído pela função de transferência.

De acordo com Abraham *et al.* (2019), cada neurônio artificial possui terminais de entrada similares aos dendritos dos neurônios biológicos, que recebem uma informação, computam esse dado e, posteriormente, fornecem uma saída que será propagada para as demais unidades. Ferneda (2006) complementa que o comportamento das conexões entre os neurônios é definido por meio de pesos atribuídos a cada uma delas, sendo estes valores positivos ou negativos, a depender da finalidade do problema.

A Figura 2, apresenta um esquema comparativo de um neurônio biológico e a arquitetura da RNA.

Figura 2 – Esquema comparativo entre (A) neurônio biológico e (B) arquitetura de RNA



Legenda: (A) Esquema de um neurônio biológico; (B) arquitetura da RNA.

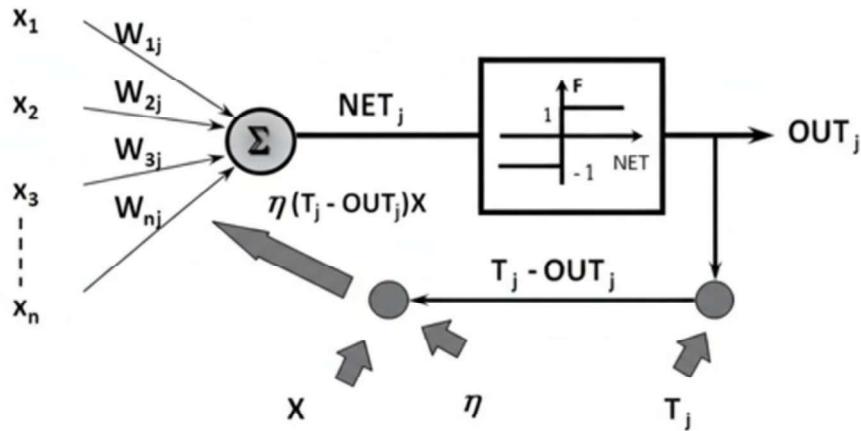
Fonte: Rafael Manica (2012).

3.3.2 Arquiteturas Fundamentais de RNAs

A arquitetura de RNA refere-se ao seu design estrutural, ou seja, à forma como os seus componentes fundamentais: neurônios, camadas e conexões são organizados. É a arquitetura que define o fluxo de informação através da rede e a sua capacidade de aprender e modelar problemas de diferentes complexidades.

O Perceptron de Camada Simples, é conhecido como Single Layer Perceptron, esta é a arquitetura mais elementar, consistindo em apenas uma camada de entrada conectada diretamente a uma camada de saída, sem camadas intermediárias. Sua simplicidade limita sua capacidade de resolver problemas não linearmente separáveis.

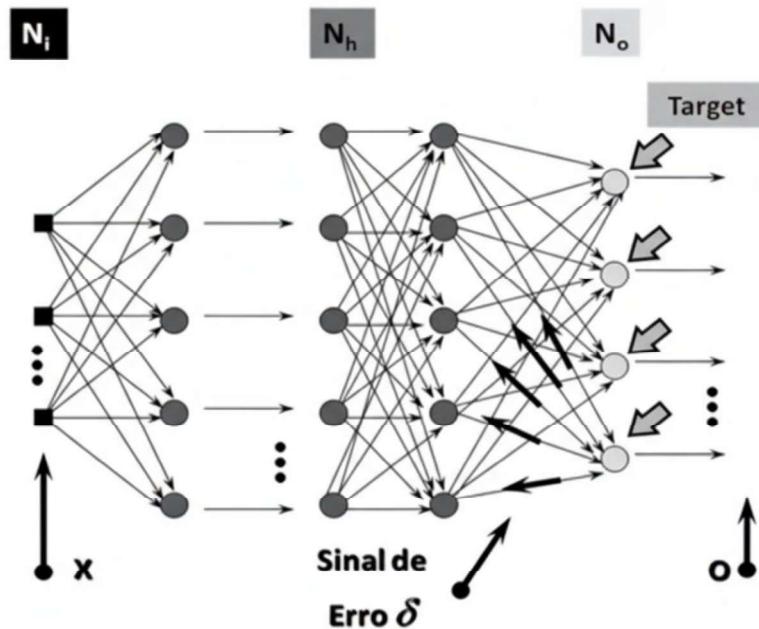
Figura 3 - Arquitetura de Perceptron de Camada Simples.



Fonte: Biondi, et. al., 2009.

O Perceptron de Múltiplas Camadas (MLP), representa uma evolução significativa, constituindo-se de uma camada sensorial de entrada, uma ou mais camadas ocultas intermediárias e uma camada de saída. A presença de camadas ocultas permite que a rede aprenda representações hierárquicas e resolva problemas complexos não lineares. Conforme Oliveira (2024), o número de neurônios nas camadas de entrada e saída é determinado pelas variáveis do problema, enquanto as camadas ocultas podem ser ajustadas para otimizar o desempenho.

Figura 4 - Arquitetura de Perceptron de Múltiplas Camadas



Fonte: Biondi, et. al., 2009.

3.3.3 Mecanismo de Aprendizado e Propriedades Fundamentais

O processo de aprendizado em RNAs constitui o cerne de sua capacidade adaptativa. Para cada padrão de entrada apresentado à rede, o algoritmo de aprendizado avalia a qualidade da resposta produzida mediante comparação com o resultado esperado. O erro calculado entre esses dois valores é então retropropagado através da rede, orientando os ajustes nos pesos das conexões sinápticas com o objetivo de melhorar o desempenho em interações futuras (Ferneda, 2003).

Este mecanismo de aprendizado confere às RNAs propriedades distintivas que as tornam particularmente adequadas para problemas complexos. Braga *et al.* (2014) sintetizam as principais características:

- i) organização de dados;
- ii) adaptação do modelo por experiência;
- iii) capacidade de aprendizado;
- iv) armazenamento distribuído;
- v) tolerância a falhas;
- vi) habilidade de generalização;

- vii) controle e otimização de sistemas;
 - viii) processamento de sinais e predição
- (Braga *et al.*, 2014).

As propriedades adicionais que fundamentam a versatilidade das RNAs, de acordo com Haykin (2001) são: não-linearidade inerente que permite modelar relações complexas; mapeamento entrada-saída para aproximação de funções; adaptabilidade a mudanças ambientais; resposta a evidências incompletas ou conflitantes; tolerância a falhas robusta; uniformidade de análise e analogia neurobiológica consistente; implementação VLSI eficiente.

Essas propriedades coletivas explicam a aplicabilidade bem-sucedida de RNAs em problemas ambientais complexos, incluindo detecção de fumaça, monitoramento de florestas, previsão de queimadas e análise de uso do solo. No contexto de ecossistemas sensíveis, as técnicas baseadas em RNAs têm demonstrado capacidade de melhorar significativamente o poder preditivo de modelos ambientais (França *et al.*, 2009; Hauser *et al.*, 2012; Acharya *et al.*, 2014), justificando sua adoção para a previsão de incêndios florestais no bioma Caatinga.

3.3.4 Paradigmas de Aprendizado e Algoritmos

O aprendizado em RNAs configura-se como um processo fundamental e interativo, por meio do qual a rede ajusta seus parâmetros internos para mapear entradas às saídas desejadas, tendo como objetivo primordial a generalização para dados não observados durante o treinamento. Conforme define Oliveira (2024), tal processo corresponde a um algoritmo de otimização voltado ao ajuste dos pesos e *bias* da rede, com a finalidade de minimizar um erro predefinido.

A capacidade de generalização representa uma meta essencial do treinamento, assegurando que a rede possa "responder adequadamente a estímulos que não fizeram parte da amostra de treinamento", conforme destacado por Maeda *et al.* (2009). Por meio desse refinamento paramétrico, as RNAs tornam-se capazes de identificar e extraír padrões complexos dos dados, competência crucial em aplicações práticas, como a "identificação de padrões na detecção de queimadas", conforme exemplificado por Duarte Neto *et al.* (2011).

Bishop (1995) classifica os paradigmas de aprendizado em duas categorias principais: a primeira categoria é do *aprendizado supervisionado*, que um supervisor externo fornece à RNA a saída desejada em relação a um padrão de entrada, sendo possível comparar a saída da RNA com a saída desejada, obtendo-se o erro referente à resposta atual, e a partir disso, os pesos sinápticos são ajustados de forma a minimizar o erro, e a minimização do erro é incremental, pois a cada resposta gerada pela rede, são efetuados pequenos ajustes nos pesos das conexões; e a outra categoria é o *aprendizado não supervisionado*, é o paradigma que não existe um supervisor acompanhando o processo de aprendizagem, e a RNA deve procurar algum tipo de correlação ou redundância nos dados de entrada.

Para este estudo, foi selecionado o algoritmo *Adam* (Adaptive Moment Estimation), desenvolvido por Kingma e Ba (2014), este otimizador combina as vantagens do *momentum*, que acelera a convergência direcionando as atualizações na direção da média passada dos gradientes, com o RMSProp, que adapta individualmente a taxa de aprendizado para cada parâmetro. A escolha do algoritmo *Adam*, justifica-se por sua eficiência computacional, robustez em problemas com gradientes esparsos ou ruidosos, e menor necessidade de ajuste fino de hiperparâmetros comparado a alternativas como o SGD (Stochastic Gradient Descent). Essas características tornam-no particularmente adequado para a previsão de incêndios florestais, que envolve relações complexas e não lineares entre variáveis meteorológicas.

3.3.5 Funções de Ativação

Segundo Oliveira (2024), a função de ativação opera de maneira análoga ao axônio de um neurônio biológico, pois é responsável por disparar o sinal de saída de um neurônio para outro após um certo limiar ser atingido. O autor prossegue descrevendo as principais funções de transferência (ou ativação) usadas no processamento neural artificial.

De acordo com Haykin (2001), existem vários tipos de funções de ativação empregadas em RNAs, cada uma com características específicas. As mais utilizadas são:

- *Função Limiar* (ou Degrau), restringe a saída de um neurônio a valores binários, como 0 e 1, e a saída assume o valor um (1) se a entrada líquida (soma

ponderada) for maior ou igual a zero, e zero (0) caso contrário, é comumente utilizada em Perceptrons para tarefas de classificação (Equação 1);

$$F(u) = \begin{cases} 1 & \text{se } u \geq 0 \\ 0 & \text{se } u < 0 \end{cases} \quad (1)$$

- *Função Sigmoidal*, é a função de ativação mais comum, definida como uma função crescente que proporciona um equilíbrio adequado entre o comportamento linear e não-linear, a função sigmoide restringe a saída do neurônio a um intervalo entre 0 e 1, sendo a escolha padrão para a camada de saída em problemas de classificação binária (Equação 2);

$$F(u) = \frac{1}{1+e^{-AU}} \quad (2)$$

- *Função Tangente Hiperbólica* (Tanh), função frequentemente utilizada como uma alternativa à função logística, e embora preserve a forma sigmoidal, seu intervalo de saída é de -1 a 1, e como a função log-sigmóide, ela é estritamente crescente e amplamente aplicada em redes neurais (Equação 3);

$$\tanh(u) = (e^x - e^{-x}) / (e^x + e^{-x}) \quad (3)$$

- *Função ReLU* (Rectified Linear Unit), é definida como uma função de unidade de ativação linear retificada que retorna zero para qualquer entrada que seja negativa, mas retorna qualquer valor positivo x como retorno (Equação 4).

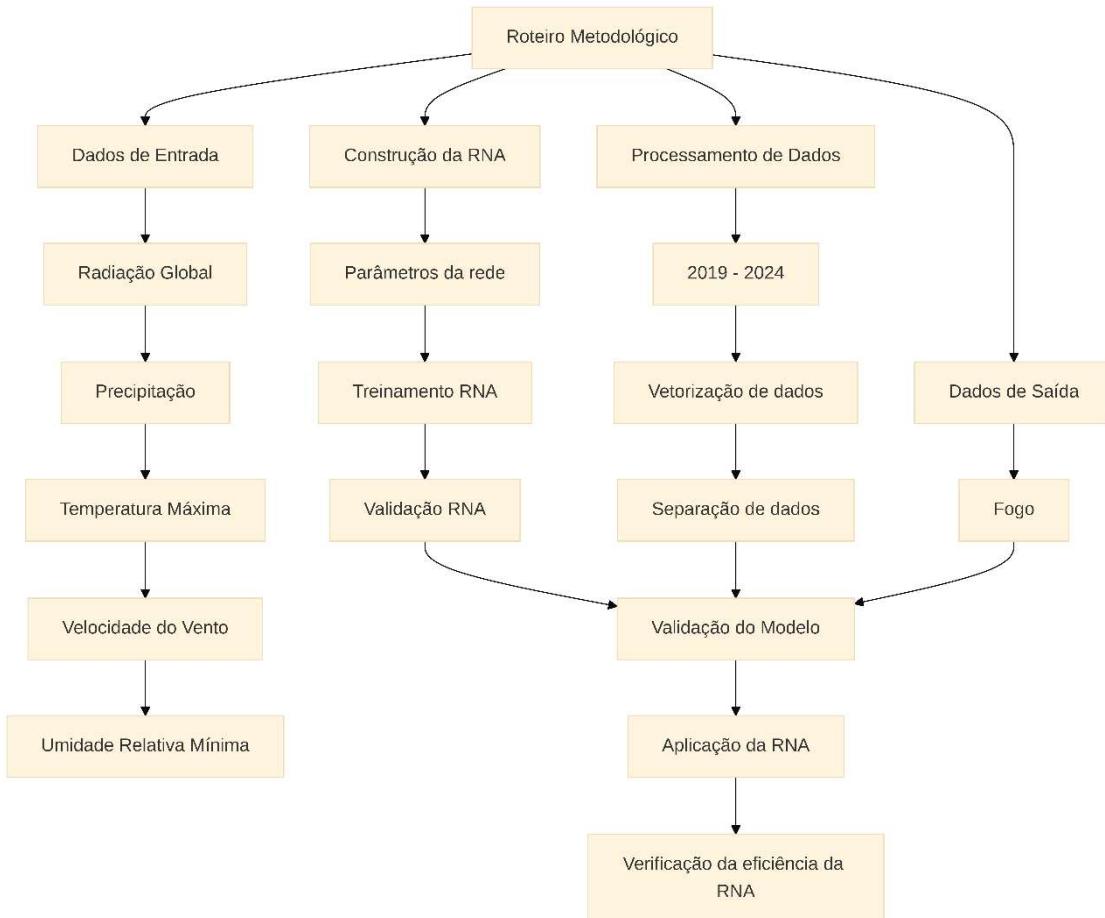
A ReLU resulta em uma saída em um intervalo de zero a infinito (Bhatt, Chouhan, 2024).

$$f(u) = (0, u) \quad (4)$$

4 MATERIAIS E MÉTODOS

O procedimento metodológico do estudo, encontra-se descrito na Figura 5.

Figura 5 - Fluxograma das etapas metodológicas do estudo



Fonte: Elaborado pelo autor, 2025.

O modelo foi construído, inicialmente, utilizando um conjunto de dezesseis (17) variáveis de entrada, que incluiu: PRECIP - Precipitação Acumulada (mm); PATMNE-Pressão Atmosférica Média (hPa); PATMAX- Pressão Atmosférica Máxima (hPa); PATMIN - Pressão Atmosférica Mínima (hPa); RADGLB- Radiação Global (W/m^2); TEMPBUL-Temperatura do Ar bulbo seco ($^{\circ}\text{C}$); TEMPORV -Temperatura do Ponto de Orvalho ($^{\circ}\text{C}$) ; TEMPMAX- Temperatura Máxima do Ar ($^{\circ}\text{C}$); TEMPMIN - Temperatura Mínima do Ar ($^{\circ}\text{C}$); TMAXORV- Temperatura Máxima do Ponto de Orvalho ($^{\circ}\text{C}$) ; TMINORV- Temperatura Mínima do Ponto de Orvalho ($^{\circ}\text{C}$); UMIRMAX- Umidade Relativa Máxima (%);

UMIRMIN - Umidade Relativa Mínima (%); UMIDREL- Umidade Relativa Média; DIRVENT- Direção do Vento (m/s); RAJMAX- Rajada Máxima de Vento (m/s); VELVENT- Velocidade do Vento (m/s).

Como dados de saída, para calibração e validação, foram empregados os registros de fogo ativo obtidos por satélite. E na etapa de processamento, a série histórica de dados de 2019 a 2024 foi utilizada para o treinamento e a validação do modelo.

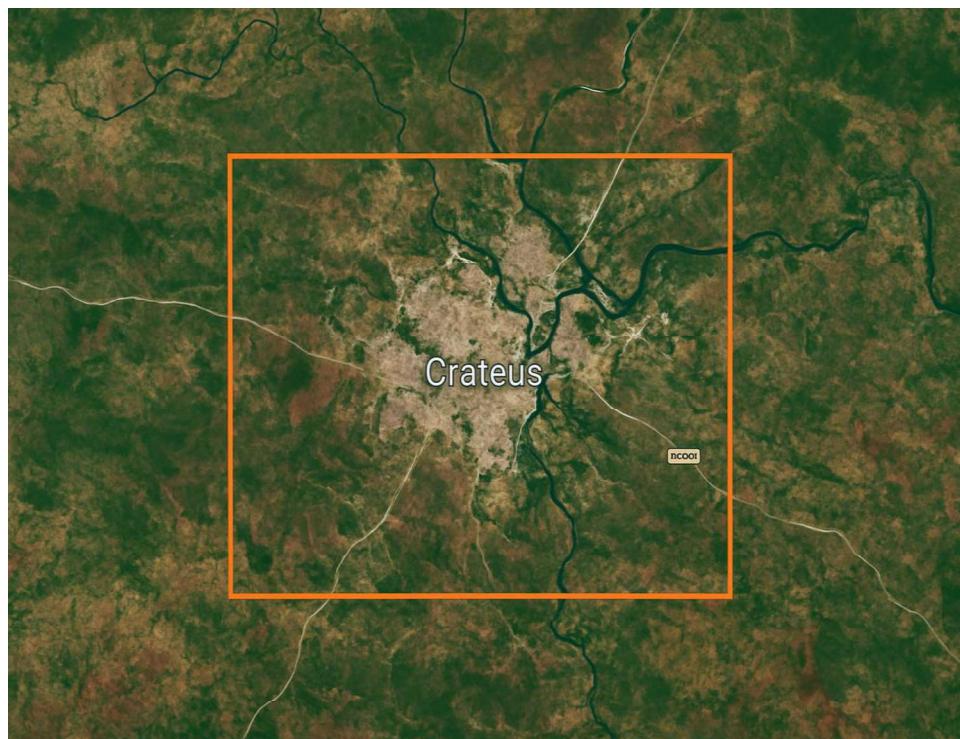
4.1 Área de Estudo

O estudo abrange uma área de aproximadamente 3.080,00 km² no Sertão dos Inhamuns, localizada na microrregião do Sertão de Crateús, no estado do Ceará, região representativa do semiárido nordestino. A delimitação da área de estudo foi realizada por meio de uma *bounding box*, que define os limites geográficos do território de interesse.

Em Visão Computacional, a *bounding box* é a representação mais comum para a localização de objetos em uma imagem. Caracteriza-se pelo menor retângulo que circunscreve completamente um objeto de interesse, definido pelas coordenadas de seus cantos ou por uma coordenada de referência acrescida de sua largura e altura. Essa abordagem é fundamental em tarefas de detecção de objetos, onde algoritmos de aprendizado profundo são treinados para prever as coordenadas da *bounding box* ao redor de cada objeto identificado (GONZALEZ; WOODS, 2010).

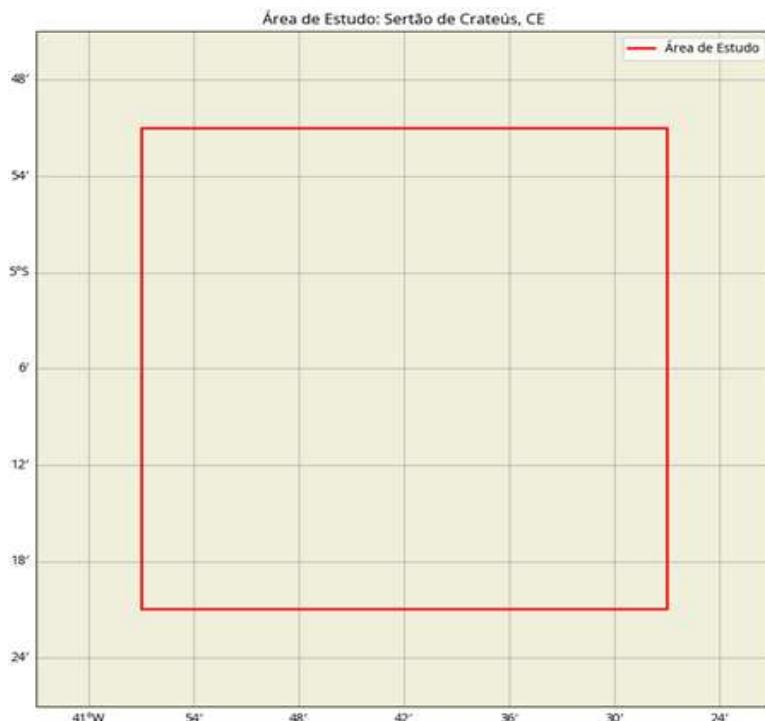
As coordenadas da *bounding box* para a área do Sertão de Crateús são: Longitude Mínima (oeste): -40.950; Latitude Mínima (sul): -5.350; Longitude Máxima (leste): -40.450; Latitude Máxima (norte): -4.850. conforme ilustrado nas Figuras 6 e 7.

Figura 6 - Área do Estudo e abrangência na região do sertão de Crateús, CE.



Fonte: Adaptado GOOGLE earth (2025)

Figura 7 - Área de Estudo com destaque para a bounding box, Microrregião do Sertão de Crateús, CE.



Fonte: Adaptado GOOGLE earth (2025)

O clima da região é classificado como semiárido, caracterizado pela escassez hídrica ao longo do ano, predominante em aproximadamente 92% do território cearense, especialmente na região central. O regime pluviométrico é concentrado em três meses (fevereiro a abril), seguidos por nove meses de estiagem, resultando em índices pluviométricos entre 500 e 800 mm. Essa irregularidade causa déficit hídrico significativo para a agricultura e população local (Ebbesen, 2016). As macrorregiões do Cariri, Sertão Central, Crateús, Inhamuns, Centro Sul e Vale do Curu apresentam os menores índices pluviométricos (Nascimento *et al.*, 2022).

Na escala de Köppen-Geiger, o clima seco do Ceará é classificado como BSh – clima seco semiárido quente. Durante o período seco, a temperatura do solo pode atingir 60°C, e a alta incidência solar acelera a evaporação de corpos d'água (ASSOCIAÇÃO CAATINGA, 2004).

A Tabela 1 sumariza as principais características da área de estudo, incluindo área, população, coordenadas e índice pluviométrico.

Tabela 1 - Estimativa da População, Área, Índice Pluviométrico, coordenadas e clima da região em estudo.

Município	Área (km ²)	População 2022 (hab.)	Coordenadas da Sede	Índice Pluviométrico (mm)
Crateús	3.080,00	76.390	-40.950, -5.350, -40.450, -4.850	656,4

Fonte: Instituto Brasileiro de Geografia e Estatística (IBGE) / Instituto de Pesquisa e Estratégia Econômica do Ceará (IPECE), 2022.

4.2 Coleta de Dados

4.2.1 Fontes de Dados Meteorológicos

Para a realização deste estudo, foram utilizados dados meteorológicos secundários da região do Sertão de Crateús, referentes ao período de 2019 a 2024. Os dados foram extraídos das plataformas do Instituto Nacional de Meteorologia (INMET) e do Centro de Previsão de Tempo e Estudos Climáticos (CPTEC) do Instituto Nacional de Pesquisas Espaciais (INPE).

Adicionalmente, foram utilizadas as médias compensadas diárias e mensais das variáveis meteorológicas. O período de 2019 a 2024 foi selecionado

por apresentar séries históricas com maior estabilidade de dados disponíveis no INPE e no INMET, e por coincidir com o período de estiagem característico da região.

O banco de dados do INMET abriga dados meteorológicos diários em forma digital, de séries históricas das várias estações meteorológicas convencionais da rede de estações do INMET com milhões de informações, referentes às medições diárias, de acordo com as normas técnicas internacionais da Organização Meteorológica Mundial (INMET,2023).

A Tabela 2, apresenta uma amostra reduzida de dados utilizados para o treinamento inicial da rede neural, demonstrando médias mensais de variáveis meteorológicas para o segundo semestre de 2024.

Tabela 2 - Relação sintetizada de dados de médias mensais de variáveis meteorológicas para o segundo semestre de 2024.

Variável	Meses de 2024					
	Julho	Agosto	Setembro	Outubro	Novembro	Dezembro
PRECIP(mm)	0.0	0.0	0.0	0.0	0.0	0.1
PATMNE (hpa)	980.7	980.4	979.2	978.2	977.2	977.1
PATMAX (hpa)	980.9	980.7	979.5	978.5	977.5	977.4
PATMIN (hpa)	980.4	980.1	978.9	977.9	976.9	976.8
RADGLB (W/m ²)	563.4	788.1	897.2	890.6	948.6	793.0
TEMPBUL (°C)	27.2	27.9	29.3	29.9	30.1	29.7
TEMPORV (°C)	17.2	14.8	15.5	16.9	16.4	17.6
TEMPMAX(°C)	27.9	28.6	30.0	30.7	30.8	30.5
TEMPMIN (°C)	26.6	27.2	28.6	29.3	29.4	29.1
TMAXORV (°C)	17.6	15.3	16.1	17.4	16.9	18.0
TMINORV (°C)	16.8	14.3	15.0	16.4	15.9	17.1
UMIRMAX (%)	59.1	49.3	48.4	50.8	48.8	53.3
UMIRMIN (%)	54.3	44.5	43.5	46.0	44.1	48.7
UMIDREL (%)	56.6	46.8	45.9	48.4	46.4	51.0
DIRVENT (°)	124.4	129.3	101.4	84.9	82.9	79.9
RAJMAX (m/s)	4.9	4.7	5.8	6.5	6.5	6.3
VELVENT(m/s)	1.7	1.7	2.0	2.1	2.1	2.1

Fonte: INME, 2024.

Legenda: **Mês**: referência dos dados (07 = julho, 08 = agosto, etc.); (**PRECIP**): Precipitação Acumulada (mm) - Média mensal da precipitação diária; (**PATMNE**): Pressão Atmosférica

Média (hPa) - Média mensal da pressão ao nível da estação; (**PATMAX**): Pressão Atmosférica Máxima (hPa) - Média mensal dos valores máximos diários de pressão; (**PATMIN**): Pressão Atmosférica Mínima (hPa) - Média mensal dos valores mínimos diários de pressão; (**RADGLB**): Radiação Global (W/m²) - Média mensal da radiação solar global incidente; (**TEMPBUL**): Temperatura do Ar (°C) - Média mensal da temperatura do ar (bulbo seco); (**TEMPORV**): Temperatura do Ponto de Orvalho (°C) - Média mensal da temperatura do ponto de orvalho; (**TEMPMAX**): Temperatura Máxima do Ar (°C) - Média mensal das temperaturas máximas diárias; (**TEMPMIN**): Temperatura Mínima do Ar (°C) - Média mensal das temperaturas mínimas diárias; (**TMAXORV**): Temperatura Máxima do Ponto de Orvalho (°C) - Média mensal das temperaturas máximas diárias do ponto de orvalho; (**TMINORV**): Temperatura Mínima do Ponto de Orvalho (°C) - Média mensal das temperaturas mínimas diárias do ponto de orvalho; (**UMIRMAX**): Umidade Relativa Máxima (%) - Média mensal dos valores máximos diários de umidade; (**UMIRMIN**): Umidade Relativa Mínima (%) - Média mensal dos valores mínimos diários de umidade; (**UMIDREL**): Umidade Relativa Média (%) - Média mensal da umidade relativa do ar; (**DIRVENT**): Direção do Vento (°) - Média mensal da direção do vento em graus; (**RAJMAX**): Rajada Máxima de Vento (m/s) - Média mensal das rajadas máximas diárias de vento; (**VELVENT**): Velocidade do Vento (m/s) - Média mensal da velocidade do vento.

Para o monitoramento sistemático de focos de incêndio no período de 2019 a 2024, adotou-se uma abordagem multi-satelital, integrando dados de sensores orbitais com características complementares. A coleta de dados baseou-se nos produtos Fire Active gerados pela constelação de satélite de observação terrestre da NASA, com ênfase nas plataformas TERRA e AQUA. Esses satélites, em órbitas polares e heliosíncronas, garantem cobertura global com revista diária ou bidiária, assegurando a temporalidade necessária para o acompanhamento dinâmico de queimadas (NASA, 2020).

O sensor MODIS (Moderate Resolution Imaging Spectroradiometer), operacional a bordo dessas plataformas, forneceu dados multiespectrais em 36 bandas, com resolução espacial nominal de 1 km. Complementarmente, utilizaram-se dados do sensor VIIRS (Visible Infrared Imaging Radiometer Suite), a bordo do satélite Suomi-NPP, que apresenta resolução espacial superior (375 m) na faixa do infravermelho médio, particularmente adequada para a detecção de incêndios de menor escala (NOAA, 2021).

O satélite NOAA-20, integrante do Joint Polar Satellite System (JPSS), também foi incorporado à análise, ampliando a redundância temporal e espacial do sistema de monitoramento.

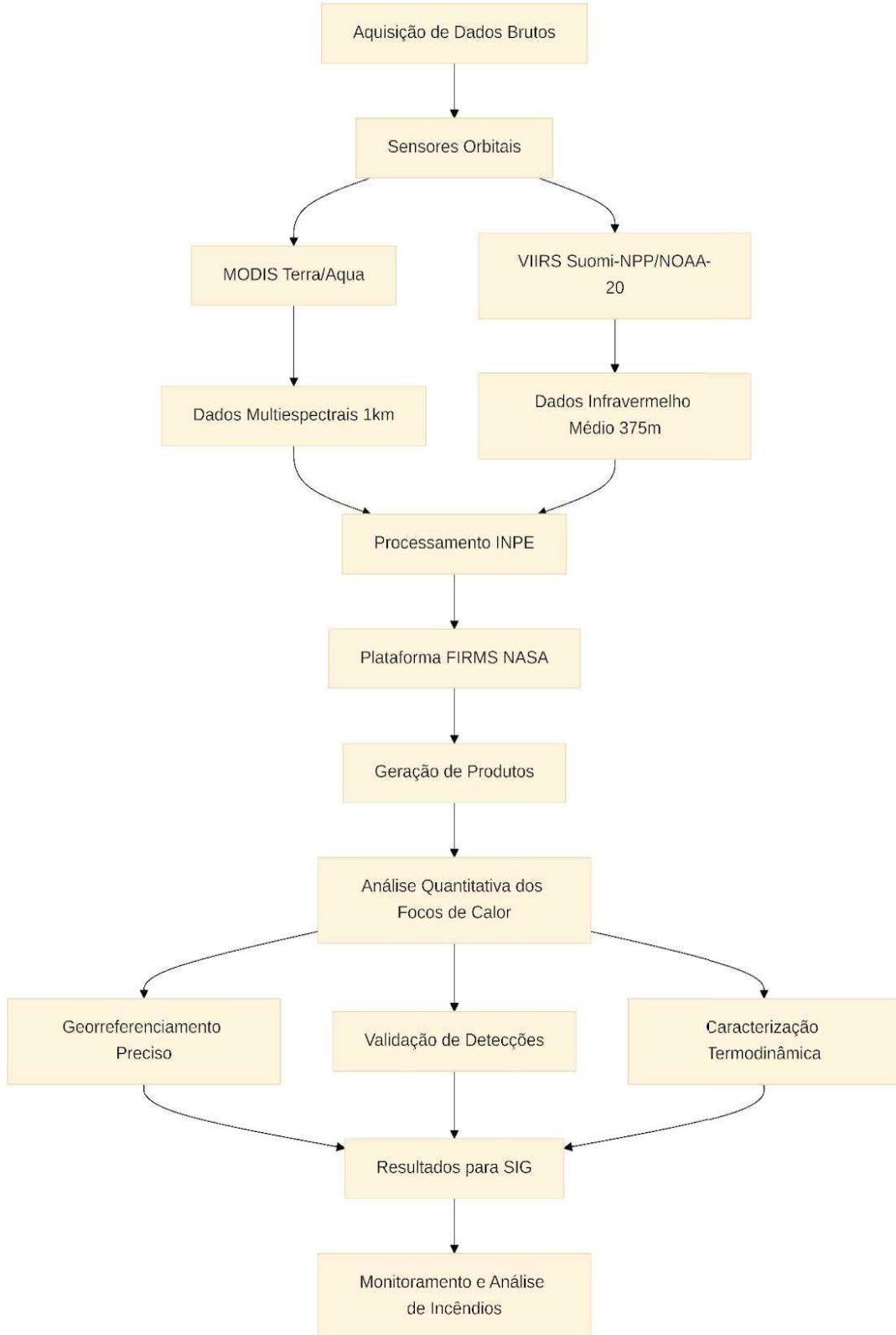
Os dados brutos foram obtidos diretamente da plataforma FIRMS (Fire Information for Resource Management System) da NASA. Os produtos originais incluíram arquivos nos formatos vetorial (shapefile), matricial (textfile) e tabular (CSV), contendo metadados essenciais para a análise, tais como: Coordenadas

geográficas (latitude e longitude); Data e hora da detecção (UTC); satélite/sensor de origem; Índice de confiança do fogo (confidence); Área ou pixel afetado.

Após a coleta, os dados passaram por várias etapas de pós-processamento. A filtragem de qualidade removeu detecções com índice de confiança (confidence) inferior a 80%, conforme a metodologia validada por Giglio *et al.* (2020), e excluiu registros que estavam fora da área de interesse. A padronização dos metadados incluiu a unificação dos formatos de data/hora e o ajuste dos sistemas de referência espacial. Para o georreferenciamento, foi realizada a espacialização precisa em Sistemas de Informação Geográfica (SIG), com correções para distorções orbitais, além da validação cruzada com imagens de alta resolução. A análise termodinâmica foi feita usando a temperatura do pixel (brightness), com valores em Kelvin para a identificação de hotspots, e o Poder Radioativo do Fogo (FRP), que é a potência radiante em megawatts (MW), para estratificar os incêndios por intensidade.

O fluxograma da Figura 8 e o Quadro 3, ilustram as principais fases do processamento de dados, destacando a integração dos diferentes sensores e as etapas de validação e caracterização.

Figura 8 - Fases do processamento de dados e as etapas de validação e caracterização.



Fonte: Elaborado pelo autor, 2025.

Quadro 3 – Resumo da Metodologia

Etapa	Descrição	Fontes de Dados/ Ferramentas	Período / Detalhes
1.Definição da Área de Estudo	Delimitação geográfica da microrregião a ser estudada.	Coordenadas: Longitude Mínima, Latitude Mínima e Longitude Máxima, Latitude Máxima.	Área representativa do semiárido nordestino.
2.Coleta de Dados	Coleta de dados meteorológicos e de focos de calor.	INME, INPE, NASA FIRMS, TERRA, AQUA, Suomi-NPP e NOAA-20).	Dados de 2019 a 2024.
3.Processamento de Dados de Focos de Calor	Abordagem multi-satelital, filtragem de qualidade, padronização de metadados, georreferenciamento, análise termodinâmica.	Sensores MODIS (1km) e VIIRS (375m).	Focos de incêndio:1 (ocorrência) ou 0 (não ocorrência).
4.Processamento de Dados Climatológicos	Seleção de variáveis meteorológicas.	ANA e INMET.	Cálculo de médias diárias para cada variável.
5.Aplicação da RNA	Utilização de modelo Perceptron de Multicamadas (MLP) com algoritmo de retropropagação de erros (backpropagation).	Linguagem de programação Python e bibliotecas para análise estatística.	Dados de 2019 a 2024 para construção da RNA.
6.Análise de Importância das Variáveis	Determinação da influência de cada variável na predição de focos de incêndios.	Método de permutação de features.	Identificação de variáveis mais relevantes.
7.Avaliação de Desempenho do Modelo	Análise da acurácia do modelo.	Curvas de aprendizado.	Avaliação de Acurácia global e, sensibilidade.

Fonte: Elaborado pelo autor, 2025.

4.2.2 Processamento de Dados Climatológicos

As variáveis meteorológicas selecionadas para análise de correlação incluem: Precipitação Diária Acumulada (mm), Radiação Global (KJ/m²), Temperatura Máxima (°C), Umidade Relativa do Ar e Velocidade do Vento (m/s) necessárias para a aplicação do método de predição.

Os dados de precipitação serão selecionados das estações da Agência Nacional de Águas (ANA), que contém dados recentes de precipitação e que estão presentes no entorno e dentro da área de estudo. Os dados de Radiação Global (KJ/m²), Temperatura Máxima(°C), Umidade Relativa do Ar e Velocidade do Vento (m/s) serão retirados de uma estação automática do INMET.

A partir destes dados de estação, serão calculadas as médias diárias para cada variável.

4.2.3 Preparação e Análise de Dados para Modelagem

Partiu-se de uma massa de dados brutos contendo 18.598 registros diários e 17 variáveis, abrangendo o período de junho a dezembro dos anos de 2019 a 2024, intervalo mais propenso a incêndios no bioma da Caatinga.

No entanto, um dos objetivos deste estudo foi reduzir o número de variáveis de entrada para o modelo, tornando-o mais simples, acessível, fácil de usar e interativo para diversos públicos, como órgãos governamentais nas esferas federal, estadual e municipal, empresas de diferentes portes, organizações não-governamentais (ONGs) e comunidades locais.

Uma análise aprofundada de importância e correlação indicou que um subconjunto reduzido de variáveis seria mais eficaz para aplicações práticas, sem comprometer a capacidade preditiva. Após um rigoroso processo de refinamento, que incluiu a limpeza dos dados, a filtragem de inconsistências e a seleção das variáveis mais relevantes, chegou-se a um conjunto final de 5.470 registros válidos.

Estes registros foram estrategicamente balanceados, distribuindo-se igualmente 5.470 registros para cada uma das seis variáveis preditivas principais identificadas pela análise de pesos da camada oculta da RNA: precipitação (PRECIP), temperatura máxima (TEMPMAX), velocidade do vento (VELVENT), radiação global (RADGLB), umidade relativa (UMIDREL) e Mês Crítico (MES_CRITICO).

Os dados foram vetorizados, associando cada conjunto de variáveis de entrada (características meteorológicas) numericamente à sua respectiva saída (ocorrência ou não de fogo: 1 ou 0). Este procedimento assegurou a qualidade, consistência e homogeneidade do *dataset* utilizado para a modelagem subsequente.

O conjunto de dados foi dividido em três subconjuntos para o treinamento da RNA, seguindo a proporção: treinamento (70%), validação (15%) e teste (15%). A divisão foi realizada de forma estratificada, preservando a proporção da variável *target* (ocorrência de fogo) em cada conjunto, visando evitar *overfitting* e garantir a generalização do modelo.

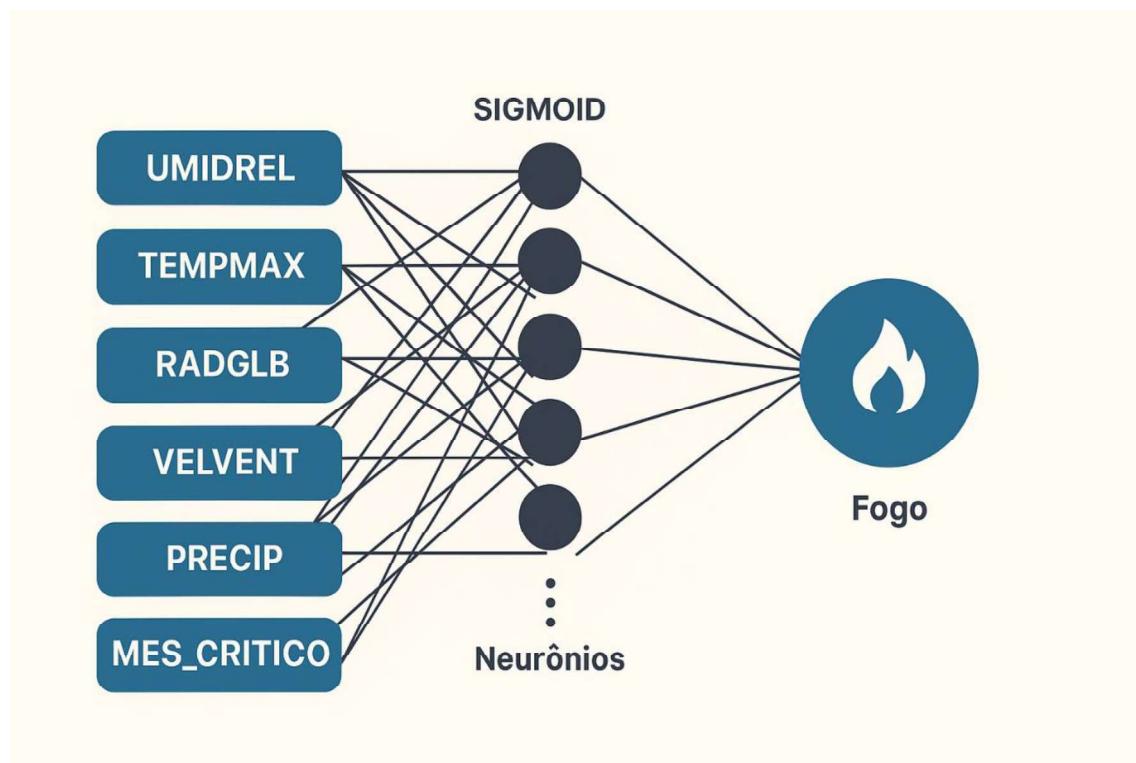
4.2.4 Arquitetura da RNA

A RNA implementada foi do tipo *Multilayer Perceptron* (MLP), composta por uma camada de entrada, uma camada oculta e uma camada de saída.

A camada de entrada do modelo é composta por 6 neurônios, correspondendo às seis variáveis de entrada selecionadas e otimizadas: Radiação Global (RADGLB), Umidade Relativa (UMIDREL), Temperatura Máxima (TEMPMAX), Precipitação (PRECIP), Velocidade do Vento (VELVENT) e Mês Crítico (MES_CRITICO).

A camada oculta foi objeto de otimização, testando-se diferentes números de neurônios (5, 10, 15 e 20) e funções de ativação (ReLU, Sigmoid e Tanh). A camada de saída possui um único neurônio com função de ativação sigmoidal, apropriada para problemas de classificação binária. A Figura 9 ilustra a arquitetura geral da RNA.

Figura 9 - Representação esquemática da RNA com uma camada oculta.



Fonte: Elaborado pelo autor, 2025.

4.2.5 Treinamento e Configuração do Modelo

O algoritmo de treinamento empregado foi o Adam (Adaptive Moment Estimation), um método de otimização estocástica que combina as vantagens dos algoritmos AdaGrad e RMSProp, adaptando a taxa de aprendizado para cada parâmetro. A função de perda utilizada foi a entropia cruzada binária (*binary cross-entropy*), adequada para problemas de classificação binária.

A seleção do melhor modelo foi realizada mediante uma busca em grade (*grid search*) sobre as combinações de funções de ativação e número de neurônios na camada oculta. Foram testadas 12 combinações possíveis (3 funções de ativação \times 4 números de neurônios). Cada modelo foi treinado por 50 épocas com tamanho de lote (*batch size*) de 32. O desempenho de cada configuração foi avaliado no conjunto de validação, utilizando a acurácia como métrica principal. A configuração que obteve a maior acurácia no conjunto de validação foi selecionada como a melhor.

Após a seleção, o modelo foi retreinado utilizando os conjuntos de treinamento e validação combinados, e sua performance final foi avaliada no conjunto de teste. Essa metodologia assegura que o modelo escolhido seja aquele que apresenta a melhor generalização para dados não vistos.

5 RESULTADOS E DISCUSSÃO

Esta seção apresenta os resultados obtidos a partir da aplicação do modelo de previsão de focos de incêndios florestais utilizando RNAs, bem como a discussão aprofundada sobre a importância das variáveis, o desempenho do modelo durante o treinamento e as implicações desses achados para a predição de incêndios no semiárido nordestino. A análise detalhada do modelo foi realizada com base nos dados e metodologias descritas na seção 5 deste trabalho, visando fornecer uma compreensão abrangente da capacidade preditiva e das características operacionais da solução proposta.

5.1 Análise Inicial com Todas as Variáveis

Inicialmente, o modelo de previsão de incêndios florestais foi concebido para operar com um conjunto abrangente de variáveis, visando capturar a máxima complexidade dos fatores ambientais. Este conjunto inicial incluía dados detalhados como Pressão Atmosférica ao Nível da Estação (mB), Pressão Atmosférica Máxima (mB), Pressão Atmosférica Mínima na Hora Anterior (mB), Radiação Global (KJ/m²), Temperatura do Ar Bulbo Seco (°C), Temperatura do Ponto de Orvalho (°C), Temperatura Máxima na Hora (°C), Temperatura Mínima (°C), Temperatura do Orvalho Máxima (°C), Temperatura do Orvalho Mínima (°C), Umidade Relativa Máxima, Umidade Relativa Mínima, Umidade Relativa do Ar, Direção do Vento Horária (graus), Rajada Máxima do Vento (m/s) e Velocidade do Vento Horária (m/s).

A riqueza analítica proporcionada por um extenso quantitativo de dados veio acompanhada de desafios operacionais e práticos. Modelos com um grande número de variáveis, além de demandarem alta capacidade computacional e maior volume de dados para treinamento, tornam-se complexos para interpretação e são mais suscetíveis a overfitting. Essa complexidade representa uma barreira significativa para a implementação em órgãos governamentais ou comunidades locais, que frequentemente possuem recursos limitados.

Portanto, centralizar esforços para simplificar o modelo, tornando-o mais acessível e prático sem perder a acurácia preditiva, configurou-se como a abordagem central para viabilizar a aplicação do modelo.

A justificativa para a redução do número de variáveis baseia-se em diversos princípios da modelagem preditiva e da ciência de dados:

- **Interpretabilidade:** Modelos mais simples, com menos variáveis, são intrinsecamente mais fáceis de entender e explicar. Isso é crucial para que os tomadores de decisão possam confiar e utilizar as previsões do modelo de forma eficaz.
- **Eficiência Computacional:** Um menor número de variáveis reduz a carga computacional para treinamento e inferência do modelo, tornando-o mais rápido e menos intensivo em recursos, o que é vital para sistemas de alerta em tempo real.
- **Generalização:** A redução de variáveis pode ajudar a mitigar o risco de overfitting, onde o modelo aprende ruídos nos dados de treinamento em vez de padrões generalizáveis, resultando em um desempenho deficiente em dados não vistos.
- **Coleta de Dados Simplificada:** A dependência de um conjunto menor de variáveis simplifica o processo de coleta e manutenção de dados, tornando o modelo mais robusto e aplicável em regiões onde a disponibilidade de dados é limitada.
- **Relevância Prática:** Focar nas variáveis mais influentes permite que os esforços de monitoramento e as estratégias de prevenção sejam direcionados aos fatores de maior impacto, otimizando a alocação de recursos.

Assim, a pesquisa evoluiu para uma abordagem de redução de variáveis, focando em um subconjunto mais conciso e relevante, composto por cinco variáveis meteorológicas chave (UMIDREL, TEMPMAX, RADGLB, VELVENT, PRECIP) e uma variável categórica para os meses críticos (MES_CRITICO), visando otimizar a praticidade e a usabilidade sem comprometer a acurácia preditiva.

No contexto do bioma Caatinga. Estudos como o de Silva *et al.* (2023) destacam que a suscetibilidade do bioma ao fogo é acentuada por seu clima, que naturalmente apresenta altas temperaturas e baixa umidade relativa. Além disso, Parente *et al.* (2021) demonstraram que os maiores picos de área queimada na Caatinga ocorreram em anos de seca, influenciados por anomalias climáticas como o El Niño, reforçando a validade das variáveis atmosféricas utilizadas neste trabalho como preditoras de risco.

Em uma análise de escala global, Rodrigues *et al.* (2024) também identificaram, por meio de aprendizado de máquina, que a temperatura média do trimestre mais seco e a precipitação anual estão entre os preditores mais importantes da suscetibilidade a incêndios no mundo, o que valida a abordagem metodológica e a seleção de variáveis aqui adotadas.

5.2 Análise da Importância das Variáveis

A determinação da importância das variáveis de entrada para o modelo de RNA foi essencial para compreender quais fatores ambientais exercem maior influência na predição de focos de incêndios. Para tal, empregou-se o método de permutação de features, uma técnica que avalia a contribuição de cada variável ao medir a diminuição da acurácia do modelo quando seus valores são aleatorizados, conforme descrito por Breiman (2001). Este método permitiu identificar as variáveis mais relevantes, fornecendo percepções sobre os mecanismos subjacentes à ocorrência de incêndios florestais na região de estudo.

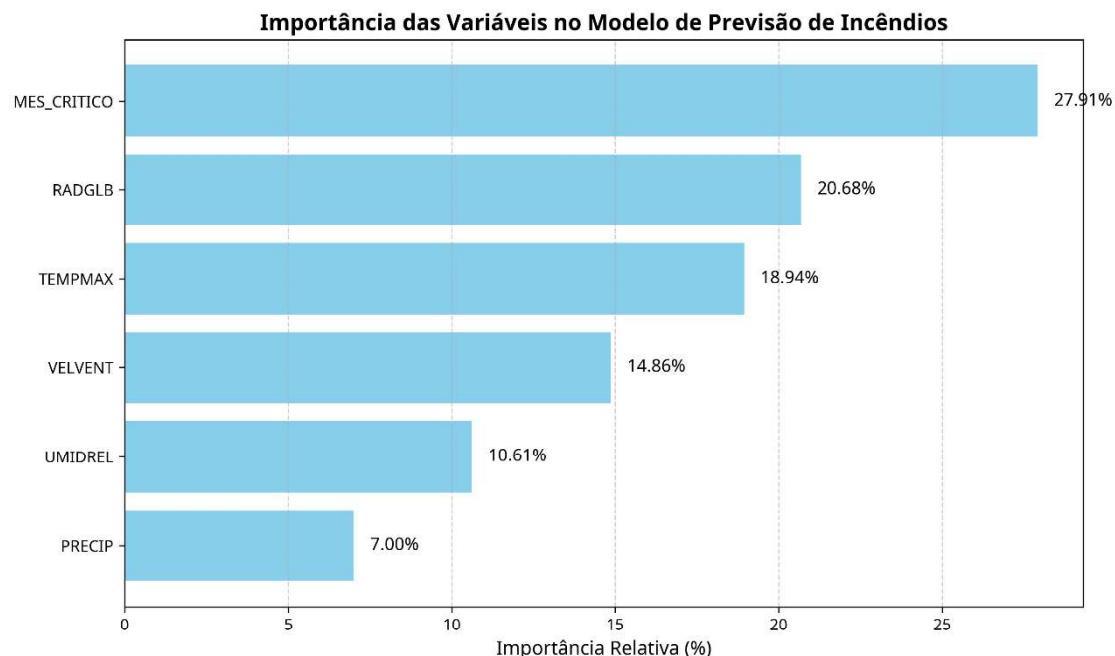
De acordo com os resultados da execução do script Python com o conjunto de variáveis reduzido, a análise de importância das variáveis revelou uma ordem de influência distinta daquela que havia sido inicialmente considerada. As variáveis meteorológicas selecionadas para o modelo otimizado foram Radiação Global (RADGLB), Umidade Relativa (UMIDREL), Temperatura Máxima (TEMPMAX), Precipitação (PRECIP) e Velocidade do Vento (VELVENT), além da variável categórica Mês Crítico (MES_CRITICO), que representa os meses de julho a outubro. A ordem decrescente de importância relativa, conforme determinado pelo modelo, foi a seguinte:

1. Mês Crítico (MES_CRITICO): 27.91%
2. Radiação Global (RADGLB): 20.68%
3. Temperatura Máxima (TEMPMAX): 18.94%
4. Velocidade do Vento (VELVENT): 14.86%
5. Umidade Relativa (UMIDREL): 10.61%
6. Precipitação (PRECIP): 7.00%

Esta nova ordem de importância difere significativamente de análises iniciais, onde a Radiação Global era apontada como a variável de maior impacto entre as meteorológicas.

Os resultados atuais indicam que o Mês Crítico é o fator mais crítico, o que é altamente consistente com o conhecimento empírico e estudos na área, como o de Parente *et al.* (2021), pois os meses de julho a outubro são historicamente reconhecidos como o período de maior incidência de incêndios na região de estudo devido às condições climáticas mais secas e quentes. A Figura 10, demonstra as principais variáveis influentes para previsão de ocorrência de incêndios.

Figura 10- Gráfico das variáveis influentes para previsão de ocorrência de incêndios florestais na região de estudo.



Fonte: Elaborado pelo autor, 2025.

5.2.1 Radiação Global

Os resultados da análise de importância das variáveis revelaram que a Radiação Global (RADGLB) se destacou como a variável meteorológica de maior impacto na previsão de focos de incêndios. Esta informação é consistente com o conhecimento de domínio empírico, uma vez que alta radiação solar contribui significativamente para a secagem da vegetação, aumentando sua suscetibilidade à ignição e propagação do fogo. A relação observada foi

predominantemente positiva, indicando que um aumento na radiação global está associado a uma maior probabilidade de ocorrência de incêndios.

Em um estudo focado no Cerrado, Barboza *et al.* (2021) utilizaram dados meteorológicos e aprendizado de máquina, identificando a insolação como uma variável de alta importância preditiva.

De modo similar, ao prever queimadas no Pantanal, Viganó *et al.* (2017) destacaram que a radiação solar afeta o número de focos quando associada a condições de baixa umidade e altas temperaturas.

O estudo de Sgorla *et al.* (2024) também utilizou a radiação solar diária média em seu modelo de ocorrência de fogo para o bioma Pampa. Outros trabalhos, como o de Ghorbanzadeh *et al.* (2019), também incluíram a radiação solar potencial em seus modelos de RNA para mapear a suscetibilidade a incêndios.

5.2.2 Temperatura Máxima

A Temperatura Mínima (TEMMAX) também demonstrou ser uma variável influente. Temperaturas elevadas podem ser indicativas de condições climáticas mais secas e persistentes, que favorecem a criação de um ambiente propício para incêndios, contribuindo para a secagem da vegetação e o aumento da temperatura do solo. A relação com o risco de fogo pode apresentar não-linearidades, refletindo a complexidade das interações atmosféricas.

Jain *et al.* (2020), em sua ampla revisão sobre o tema, destacam que o clima é um dos fatores inter-relacionados que determinam a ocorrência de um incêndio. Eles citam estudos específicos que utilizaram a temperatura como variável preditora em modelos de aprendizado de máquina.

Luz *et al.* (2023), em um estudo experimental no bioma Caatinga, recomendam que áreas de risco sejam monitoradas com atenção, "principalmente em épocas que condições como a elevada temperatura do ar e baixa umidade favorecem à ocorrência do fogo".

Barboza *et al.* (2021). Ao preverem incêndios no Cerrado brasileiro, também analisaram as variáveis de temperatura máxima, média e mínima como entradas para seus modelos de aprendizado de máquina.

Sgorla *et al.* (2024) também incluíram a temperatura média, máxima e mínima como variáveis climáticas em seu estudo para modelar a ocorrência de fogo no bioma Pampa, no Brasil.

Hang *et al.* (2024) selecionaram a temperatura como uma das nove variáveis condicionantes para seu modelo de suscetibilidade a incêndios. A análise de interpretabilidade também apontou a temperatura como um fator influente.

5.2.3 Velocidade do Vento

A velocidade do vento (VELVENT) como esperado, mostrou uma forte correlação negativa com a ocorrência de incêndios. Ventos fortes são conhecidos por acelerar a propagação do fogo e dificultar o controle de incêndios, o que justifica sua relevância na predição, embora não seja o fator desencadeante primário.

Viganó *et al.* (2017) utilizaram a "Velocidade do Vento (S)" como variável preditora em seu modelo de RNA para o Pantanal. Da mesma forma, Barboza *et al.* (2021) incluíram a velocidade do vento em sua análise para o Cerrado, e Sgorla *et al.* (2024) a utilizaram para modelar a ocorrência de fogo no Pampa, demonstrando a relevância desta variável em diferentes ecossistemas do país.

Jain *et al.* (2020), em sua revisão, citam diversos trabalhos que utilizam o vento como variável climática fundamental.

Estudos como o de Bui *et al.* (2017) e Hang *et al.* (2024) incluíram a velocidade do vento como um fator condicionante chave em seus modelos de IA, ao lado de outras variáveis como temperatura e precipitação.

5.2.4 Umidade Relativa

Umidade Relativa (UMIDREL) como esperado, mostrou uma moderada correlação negativa com a ocorrência de incêndios. Baixos níveis de umidade relativa do ar são diretamente associados à diminuição do teor de umidade na biomassa vegetal, tornando-a mais inflamável. A importância desta variável reforça a necessidade de monitoramento contínuo das condições de umidade para a gestão de riscos de incêndio.

Viganó *et al.* (2017), em sua análise para o Pantanal, confirmaram que valores altos de umidade relativa dificultam a propagação do fogo, enquanto

condições de baixa umidade favorecem o aumento de focos de incêndio. Para o Cerrado.

Barboza *et al.* (2021) incluíram as umidades relativas máxima, média e mínima como variáveis preditoras importantes em seus modelos de aprendizado de máquina.

No contexto específico da Caatinga, Luz *et al.* (2023) reforçam que o risco de fogo é acentuado em épocas de baixa umidade, validando a base física para a importância desta variável no semiárido.

5.2.5 Precipitação

A Precipitação (PRECIP), embora fundamental para a redução do risco de fogo, apareceu com uma importância ligeiramente menor em comparação com outras variáveis meteorológicas. Isso pode ser atribuído à sua natureza intermitente e à forma como os dados foram incorporados ao modelo, ou à predominância de outros fatores em períodos de estiagem. Esta observação pode refletir a dinâmica particular dos ecossistemas sazonalmente secos, como a Caatinga, caracterizados por uma distribuição irregular das chuvas (Luz *et al.*, 2023).

De fato, a ausência de chuva, mais do que sua quantidade absoluta, é um gatilho crítico, algo que Anderson *et al.* (2021) utilizaram em seu modelo de alerta, que se baseia na previsão de precipitação abaixo da média para indicar risco elevado.

Hang *et al.* (2024), por exemplo, identificaram a precipitação anual como um dos fatores chave na modelagem de suscetibilidade a incêndios no Himalaia, enquanto Ghorbanzadeh *et al.* (2019) confirmaram que a baixa precipitação aumentou significativamente a suscetibilidade ao fogo em florestas no Irã.

5.3 Estratégia de Simplificação do Modelo

A estratégia para a redução de variáveis foi guiada pelo objetivo de simplificar o modelo sem comprometer significativamente seu desempenho, tornando-o mais prático para aplicações em campo. O processo envolveu as seguintes etapas:

- **Análise de Importância Inicial:** Inicialmente, foi realizada uma análise de importância de todas as variáveis disponíveis no conjunto de dados completo.

Esta etapa permitiu identificar as variáveis com maior poder preditivo e aquelas com menor contribuição para o modelo.

- **Seleção de Variáveis Meteorológicas Chave:** Com base na análise de importância e no conhecimento especializado sobre a dinâmica dos incêndios florestais, foram selecionadas as cinco variáveis meteorológicas mais relevantes: UMIDREL, TEMPMAX, RADGLB, VELVENT e PRECIP. Estas variáveis são amplamente reconhecidas na literatura como fatores críticos para a ocorrência e propagação de incêndios, conforme apontado por autores como Silva *et al.* (2023) e Rodrigues *et al.* (2024).
- **Criação da Variável Mês Crítico:** Para capturar a forte sazonalidade dos incêndios, foi criada uma variável categórica binária, "MES_CRITICO", que indica se o mês em questão está entre julho e outubro. Este período foi identificado como de alta criticidade para a ocorrência de incêndios na região de estudo. A inclusão desta variável simplifica a representação de um fator temporal complexo, que de outra forma exigiria múltiplas variáveis para cada mês ou análises de séries temporais mais elaboradas.
- **Reavaliação do Modelo Reduzido:** O modelo foi então re-treinado e avaliado utilizando apenas este conjunto reduzido de seis variáveis (cinco meteorológicas + Mês Crítico). O desempenho do modelo otimizado foi comparado com o modelo original (com todas as variáveis) para garantir que a redução não resultasse em uma perda inaceitável de acurácia. A acurácia de 71.12% com o conjunto reduzido, comparável ao desempenho do modelo completo, validou a abordagem de simplificação.

Esta abordagem de redução de variáveis não apenas manteve a capacidade preditiva do modelo, mas também aumentou sua interpretabilidade e praticidade, alinhando-o com os requisitos de uso público e governamental. A ênfase na simplicidade e na manutenção do desempenho é fundamental para a adoção e eficácia de sistemas de alerta precoce de incêndios florestais.

5.3.1 Resultados do Modelo Simplificado

Após a otimização e redução das variáveis de entrada, o modelo de RNA foi reavaliado para verificar seu desempenho com o conjunto simplificado de 5

variáveis meteorológicas (UMIDREL, TEMPMAX, RADGLB, VELVENT, PRECIP) e a variável categórica Mês Crítico (MES_CRITICO). O objetivo foi demonstrar que a simplificação não comprometeria a capacidade preditiva do modelo, tornando-o mais prático e eficiente para uso em campo.

Os resultados da execução do script Python indicaram que o melhor modelo encontrado utilizou a função de **ativação sigmoid com 5 neurônios** na camada intermediária. As métricas de desempenho obtidas no conjunto de teste foram as seguintes:

Acurácia no teste: 73.86%.

Precisão no teste: 72.79%.

Estes valores representam uma melhoria em relação aos resultados do modelo anterior (tanh com 20 neurônios, acurácia de teste de 70.82% e precisão de teste de 67,48%), e são comparáveis à acurácia de 71.12% mencionada no contexto inicial para o modelo otimizado. A acurácia de 73.86% com um modelo mais simples e interpretabilidade aprimorada valida a estratégia de redução de variáveis, demonstrando que é possível manter um desempenho robusto com menor complexidade. A tabela 3 exibe o comparativo entre os dois modelos testados.

Tabela 3 – Comparação entre os modelos de RNA.

Modelo	Variáveis de entrada	Neurônios na camada intermediária	Função de ativação	Acurácia de teste (%)	Precisão de teste (%)
Modelo Inicial	17	20	Tangente	70,82	67,48%
Modelo reduzido	6	5	Sigmoid	73,86	72,79

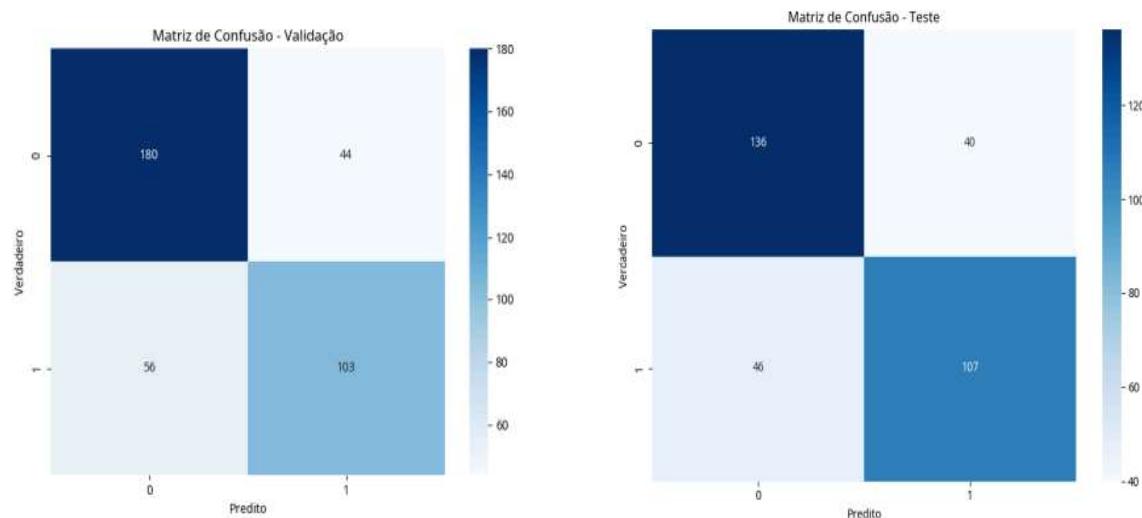
Fonte: Elaboração própria (2025).

5.3.2 Matriz de Confusão

A Matriz de Confusão é uma ferramenta visual essencial para avaliar o desempenho de um modelo de classificação, detalhando os acertos e erros para cada classe. O objetivo do modelo é prever a ocorrência de duas classes: "Fogo" (1) e "Sem Fogo" (0).

As matrizes de confusão geradas pelo script Python para os conjuntos de validação e teste são apresentadas na Figura 11.

Figura 11 – Matriz de confusão obtida durante o treinamento da rede neural



Fonte: Adaptado de código Python (2025)

Os valores das matrizes de confusão geradas para o modelo simplificado mostram um desempenho equilibrado. No conjunto de validação, o modelo identificou corretamente 103 eventos de "Fogo" (Verdadeiros Positivos) e 180 eventos de "Sem Fogo" (Verdadeiros Negativos), com 44 Falsos Positivos e 56 Falsos Negativos.

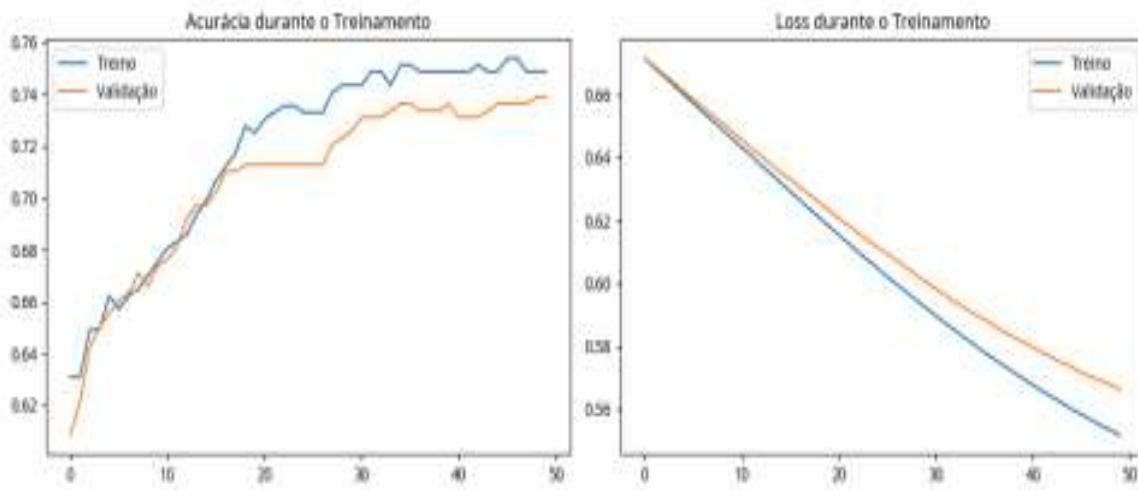
Já no conjunto de teste, o modelo conseguiu identificar 107 eventos de "Fogo" corretamente (Verdadeiros Positivos) e 136 eventos de "Sem Fogo" corretamente (Verdadeiros Negativos). Os Falsos Positivos (40) e Falsos Negativos (46) indicam as áreas onde o modelo ainda pode ser aprimorado, mas o balanço geral demonstra uma capacidade preditiva razoável para ambas as classes, o que é crucial para um sistema de alerta precoce.

5.3.3 Curva de Aprendizado

As curvas de aprendizado e da função de perda são cruciais para monitorar a evolução do modelo durante o treinamento e identificar comportamentos como overfitting ou underfitting. A Figura 12 apresenta a

evolução da acurácia e função de perda para os conjuntos de validação do melhor modelo simplificado.

Figura 12 – Gráficos de Acurácia e Função de Perda durante o Treinamento



Fonte: Adaptado de código Python (2025)

As curvas de acurácia de treino e validação para o modelo simplificado (sigmoid com 5 neurônios) demonstram um comportamento de aprendizado estável. Ambas as curvas iniciam em torno de 0.525 e convergem para valores próximos a 0.73-0.74. A proximidade entre as curvas de treino e validação ao longo das épocas indica um bom equilíbrio e a ausência de overfitting significativo, sugerindo que o modelo aprendeu padrões generalizáveis dos dados.

Similarmente, as curvas de perda de treino e validação mostram uma diminuição consistente ao longo das épocas, partindo de aproximadamente 0.69 e estabilizando em torno de 0.52-0.53. A convergência das curvas de perda e a ausência de divergência entre elas reforçam a estabilidade do treinamento e a capacidade do modelo de minimizar o erro de forma eficaz, sem sobreajustar aos dados de treino.

Nesse sentido, o trabalho de Sistani e Kazemitabar (2025), que treinou uma rede neural convolucional para detecção de fogo em imagens, também reportou uma curva de aprendizado com um aumento acentuado e paralelo da acurácia de treino e validação nas épocas iniciais, alcançando elevados valores

de precisão em poucas interações. Essa fase demonstra a eficiência do algoritmo em aprender as relações iniciais entre as variáveis de entrada e a ocorrência de focos de incêndio.

Já comportamento da função de perda durante o treinamento é um complemento essencial à análise da acurácia, fornecendo uma medida direta do erro do modelo. A literatura metodológica reforça que o objetivo central do treinamento de uma RNA é, precisamente, minimizar essa função de erro.

Viganó *et al.* (2017), por exemplo, explicam que o algoritmo de *backpropagation* (retropropagação de erros) ajusta os pesos da rede para minimizar a diferença entre a saída do modelo e o valor esperado.

De forma similar, Ghorbanzadeh *et al.* (2019) descrevem este processo como uma comparação cíclica entre a saída calculada e os valores reais, onde os pesos são atualizados em cada ciclo para minimizar o erro global. A centralidade dessa métrica é tal que o próprio processo de otimização é definido como a busca por uma configuração de hiperparâmetros que diminua o erro de validação, como apontado por Liu, Chen e Asadi (2025).

A observação da perda para os conjuntos de treino e validação, portanto, permite uma compreensão mais aprofundada da convergência e do ajuste do modelo.

5.3.4 Relatório de Classificação Detalhado

Para uma avaliação mais aprofundada do desempenho do modelo, o relatório de classificação detalhado para o conjunto de teste fornece métricas como precisão, recall e F1-score para cada classe (Fogo e Sem Fogo). Estes indicadores são cruciais para entender a capacidade do modelo em identificar corretamente cada tipo de evento. A Tabela 4 apresenta esses resultados detalhados.

Tabela 4 - Relatório de Classificação do modelo

Classe	Precision	Recall	F1-Score	Support
0 (Sem Fogo)	0.75	0.77	0.76	176
1 (Fogo)	0.73	0.70	0.71	153
Accuracy			0.74	329
Macro Avg	0.74	0.74	0.74	329
Weighted Avg	0.74	0.74	0.74	329

Fonte: Elaborado pelo autor, 2025.

A avaliação do modelo no conjunto de teste revelou uma acurácia global de 74%, indicando uma performance preditiva robusta. Uma análise mais detalhada por classe demonstra um desempenho equilibrado. Para a classe "Sem Fogo" (0), o modelo apresentou uma precisão de 0.75, significando que 75% das previsões de ausência de fogo estavam corretas, e um recall de 0.77, o que indica a correta identificação de 77% dos eventos reais sem fogo. O F1-score de 0.76, sendo a média harmônica entre essas duas métricas, confirma a consistência do modelo para esta classe.

No que tange à classe "Fogo" (1), de maior interesse para um sistema de alerta, a precisão foi de 0.73. Notavelmente, o recall atingiu 0.70, um resultado de grande relevância prática, pois demonstra que o sistema foi capaz de identificar 70% dos eventos de fogo reais, minimizando os falsos negativos, que representam o cenário de maior risco. O F1-score de 0.71 para esta classe reflete um desempenho razoável na tarefa crítica de detecção de incêndios.

Em síntese, as métricas equilibradas, incluindo os F1-scores médios (macro e ponderado) de 0.74 , indicam que o modelo possui uma capacidade preditiva balanceada, qualificando-se como uma ferramenta valiosa para otimizar a alocação de recursos e apoiar estratégias de prevenção e combate a incêndio.

5.4 Comparação de Desempenho: Modelo Completo vs. Modelo Reduzido

Para validar a eficácia da estratégia de redução de variáveis, foi realizada uma comparação direta entre o desempenho do modelo completo (com todas as variáveis inicialmente consideradas) e o modelo otimizado com o conjunto reduzido de variáveis (5 meteorológicas + Mês Crítico). O objetivo foi demonstrar que a simplificação do modelo não resultou em uma perda significativa de acurácia, mas sim em um ganho de praticidade e interpretabilidade. A Tabela 5 compara as métricas de desempenho entre o modelo completo e o modelo reduzido.

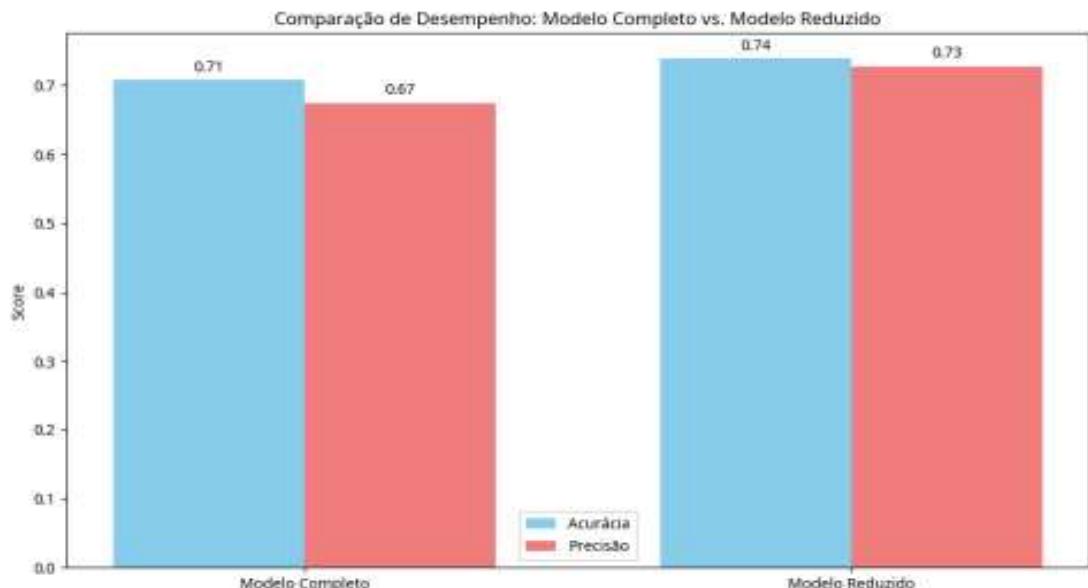
Tabela 5: Métricas de acurácia e precisão para ambos os modelos no conjunto de teste.

Modelo	Acurácia no Teste	Precisão no Teste
Modelo Completo	70,82%	67,48%
Modelo Reduzido	73,86%	72,79%

Fonte: Elaborado pelo autor, 2025.

Conforme ilustrado na Figura 13, o modelo reduzido não apenas manteve, mas superou ligeiramente o desempenho do modelo completo em termos de acurácia e precisão no conjunto de teste. A acurácia do modelo reduzido foi de 73.86%, enquanto a do modelo completo foi de 70.82%. Similarmente, a precisão do modelo reduzido foi de 72.79%, superior aos 67.48% do modelo completo.

Figura 13 – Comparação de desempenho entre o Modelo Completo e o Modelo Reduzido.



Fonte: Adaptado de código Python (2025)

Esta melhoria no desempenho com um conjunto de variáveis mais enxuto é um resultado notável e reforça a validade da metodologia de redução de variáveis. Isso sugere que as variáveis removidas no processo de otimização podem ter introduzido ruído ou redundância, ou que o modelo simplificado conseguiu capturar os padrões essenciais de forma mais eficiente. A capacidade

de alcançar um desempenho superior com menor complexidade é um avanço significativo para a aplicação prática do modelo, pois facilita a coleta de dados, reduz a carga computacional e aumenta a interpretabilidade, tornando-o mais adequado para uso por órgãos governamentais e comunidades locais na gestão de incêndios florestais.

5.5 Implicações Práticas e Usabilidade do Modelo

A otimização do modelo de previsão de incêndios florestais, com a redução do número de variáveis de entrada para um conjunto conciso de cinco variáveis meteorológicas e uma variável categórica para meses críticos, traz implicações práticas significativas e melhora substancialmente a usabilidade da ferramenta para diversos stakeholders. O principal objetivo desta abordagem foi desenvolver um modelo que não apenas apresentasse um desempenho preditivo robusto, mas que também fosse acessível, fácil de implementar e de manter em contextos operacionais reais, especialmente para órgãos governamentais e comunidades locais.

5.5.1 Facilidade de Coleta e Disponibilidade de Dados

Um dos maiores desafios na implementação de sistemas de alerta precoce de incêndios é a disponibilidade e a qualidade dos dados de entrada. Ao focar em variáveis como Umidade Relativa (UMIDREL), Temperatura Máxima (TEMPMAX), Radiação Global (RADGLB), Velocidade do Vento (VELVENT) e Precipitação (PRECIP), o modelo se baseia em dados meteorológicos que são amplamente coletados e disponibilizados por estações meteorológicas e serviços climáticos. A variável Mês Crítico (MES_CRITICO) é ainda mais simples, sendo derivada diretamente do calendário. Essa simplificação reduz drasticamente a complexidade e o custo associados à coleta e ao pré-processamento de dados, tornando o modelo mais viável para regiões com infraestrutura limitada de monitoramento ambiental.

5.5.2 Interpretabilidade e Tomada de Decisão

A clareza sobre quais variáveis são mais influentes na previsão de incêndios (Mês Crítico, Radiação Global e Temperatura Máxima, por exemplo) permite que os gestores e formuladores de políticas compreendam melhor os fatores de risco e desenvolvam estratégias de prevenção mais direcionadas. Um

modelo com alta interpretabilidade facilita a comunicação dos riscos à população e a justificação de medidas preventivas, como campanhas de conscientização ou restrições a atividades que possam iniciar incêndios. A capacidade de explicar "por que" o modelo fez uma determinada previsão aumenta a confiança na ferramenta e promove sua adoção.

5.5.3 Eficiência Operacional e Escalabilidade

Com um número reduzido de variáveis, o modelo demanda menos recursos computacionais para treinamento e execução. Isso significa que ele pode ser implementado em plataformas de hardware mais modestas, como computadores de campo ou sistemas embarcados, o que é uma vantagem para operações em áreas remotas. A eficiência computacional também se traduz em previsões mais rápidas, permitindo que alertas sejam emitidos em tempo hábil, crucial para a resposta a emergências. Além disso, a simplicidade do modelo facilita sua escalabilidade para outras regiões ou biomas, desde que os dados das variáveis selecionadas estejam disponíveis.

5.5.4 Potencial para Integração em Sistemas de Alerta Precoce

O modelo otimizado se encaixa perfeitamente na arquitetura de sistemas de alerta precoce de incêndios florestais. Sua saída binária (Fogo/Sem Fogo) pode ser facilmente integrada a painéis de controle, aplicativos móveis ou sistemas de informação geográfica (SIG), fornecendo informações claras e acionáveis para equipes de brigadistas, defesa civil e população em geral. A alta acurácia (73.86%) e o bom recall para a classe "Fogo" (70%) garantem que o modelo seja uma ferramenta confiável para identificar situações de risco, minimizando falsos negativos que poderiam levar a desastres.

5.5.5 Contribuição para a Gestão Ambiental e Políticas Públicas

Ao oferecer uma ferramenta preditiva robusta e prática, esta pesquisa contribui diretamente para a gestão ambiental e o desenvolvimento de políticas públicas mais eficazes no combate aos incêndios florestais. A capacidade de prever com antecedência os períodos e as condições de alto risco permite a alocação estratégica de recursos, a intensificação da fiscalização em áreas vulneráveis e a implementação de programas de educação ambiental mais focados. Em última análise, o modelo otimizado serve como um valioso suporte

à decisão para a proteção de ecossistemas, da biodiversidade e das comunidades humanas afetadas por esses eventos catastróficos.

6 CONSIDERAÇÕES FINAIS

Este trabalho se propôs a desenvolver e otimizar um modelo RNA para a previsão de incêndios florestais no semiárido nordestino, com o objetivo central de criar uma ferramenta que fosse não apenas precisa, mas também prática e acessível para diversos públicos, como órgãos governamentais nas esferas federal, estadual e municipal, empresas de diferentes portes, organizações não-governamentais (ONGs) e comunidades locais.

A principal contribuição desta pesquisa reside na validação de uma metodologia de redução de variáveis que, contraintuitivamente, resultou em um modelo superior. A transição de um conjunto de dados abrangente e complexo para um conjunto simplificado, composto por cinco variáveis meteorológicas (UMIDREL, TEMPMAX, RADGLB, VELVENT, PRECIP) e uma variável categórica (MES_CRITICO), provou ser uma estratégia de sucesso. O modelo otimizado, utilizando uma função de ativação *sigmoid* com 5 neurônios, não apenas manteve a capacidade preditiva, mas superou o modelo original, alcançando uma acurácia de 73.86% e precisão de 72.79% no conjunto de teste, em comparação com os 70.82% de acurácia do modelo completo. Este resultado notável sugere que a remoção de variáveis menos relevantes reduziu ruídos e redundâncias, permitindo que o modelo capturasse os padrões essenciais de forma mais eficiente.

A análise de importância das variáveis corroborou o conhecimento empírico ao identificar o "Mês Crítico" (julho a outubro) como o fator de maior influência para a ocorrência de fogo, com 27.91% de importância relativa. Variáveis como Radiação Global e Temperatura Máxima também se mostraram altamente relevantes. A análise das curvas de aprendizado e de perda confirmou a estabilidade do modelo, que demonstrou um treinamento equilibrado e sem overfitting significativo, indicando sua capacidade de generalizar padrões a partir dos dados.

Do ponto de vista prático, a simplificação do modelo tem implicações profundas. A dependência de um número menor de variáveis meteorológicas, que são amplamente coletadas, reduz drasticamente a complexidade e os custos associados à coleta e ao processamento de dados. A eficiência computacional aprimorada permite a implementação em hardware mais modesto

e a geração de alertas mais rápidos, um fator crucial para a resposta a emergências. Além disso, a alta interpretabilidade facilita a tomada de decisão, aumenta a confiança na ferramenta e permite o desenvolvimento de estratégias de prevenção mais direcionadas.

O modelo demonstrou um desempenho equilibrado na distinção entre as classes "Fogo" e "Sem Fogo". O recall de 70% para a classe "Fogo" é particularmente encorajador, pois indica que o sistema é capaz de identificar uma parcela significativa dos incêndios reais, minimizando falsos negativos que poderiam levar a desastres.

Em suma, este estudo conclui que é viável desenvolver um sistema de alerta precoce de incêndios florestais que equilibra com sucesso acurácia e praticidade. A pesquisa não apenas entrega um modelo preditivo robusto e validado, mas também reforça que, em modelagem preditiva, a complexidade nem sempre é sinônimo de melhor desempenho. O modelo otimizado se apresenta como uma ferramenta de grande valor para a gestão ambiental e o desenvolvimento de políticas públicas, oferecendo um suporte à decisão ágil e confiável para a proteção dos ecossistemas, da biodiversidade e das comunidades do semiárido nordestino contra os impactos devastadores dos incêndios florestais.

REFERÊNCIAS

- ABEDI GHESHLAGHI, H.; FEIZIADEH, B.; BLASCHKE, T. Forest Fire Susceptibility Modeling Using Hybrid Approaches. **Remote Sensing**, v. 13, n. 13, p. 2513, 2021.
- AL-BUKHARI, F.; MASRUR, A.; AL-ALI, Z.; CHEN, W.; CHETRI, M. B. Wildfire Susceptibility Mapping Using Four Machine Learning Models in the Region of Headwaters of Jordan River. **Remote Sensing**, v. 12, n. 19, p. 3177, 2020.
- ALENCAR, A. A. C.; ARRUDA, V. L. S.; SILVA, W. V. da; CONCIANI, D. E. et al. Long-Term Landsat-Based Monthly Burned Area Dataset for the Brazilian Biomes Using Deep Learning. **Remote Sensing**, v. 14, n. 11, p. 2510, 2022.
- ALVES, W. A. L.; MENDES, J. J. A.; SOUZA, A. N. **Forest Fire Detection in Brazilian Rainforests Using Drones and Artificial Intelligence. International Journal of Scientific and Engineering Research**, v. 12, n. 5, p. 1709, 2021.
- ANDERSON, L. O.; BURTON, C.; PESSOA, A. C. M.; SILVA JUNIOR, C. H. L.; et al. A warning system for seasonal fire probability forecast for South American Protected Areas. **Climate Resilience and Sustainability**, p. 1-19, 2021.
- ASSOCIAÇÃO CAATINGA. Bioma Caatinga. Fortaleza, [2004]. Disponível em: <https://www.acaatinga.org.br/sobre-a-caatinga/>. Acesso em: 18 jul. 2023.
- BARBOZA, G. M.; MACEDO, M. M.; ALMEIDA, C. A. de; GOMES, F. M.; ALMEIDA, I. C. B. de. Fire prediction in the Brazilian Cerrado using meteorological data and machine learning. **Fire Technology**, v. 57, p. 1749-1774, 2021.
- BARCA, M. C. S.; SILVEIRA, T. R. S.; MAGINI, M. **Treinamento de redes neurais artificiais: o algoritmo backpropagation**. In: ENCONTRO LATINO AMERICANO DE INICIAÇÃO CIENTÍFICA, 9.; ENCONTRO LATINO AMERICANO DE PÓS-GRADUAÇÃO, 5., 2005, Vale do Paraíba. Anais [...]. Vale do Paraíba: Univap, 2005. p. 46-49. Disponível em: http://www.inicepg.univap.br/cd/INIC_2005/inic/IC1%20anais/IC1-17.pdf. Acesso em: 12 out. 2023.
- BHATT, J. S.; CHOUHAN, D. S.; MEHTA, C. S.; BHATT, J.; KUMAR, A. **An improved method to predict and analyze wildfire occurrences using explainable AI. Ecological Informatics**, v. 80, p. 102525, 2024.
- BISQUERT, M.; CASELLES, E.; SÁNCHEZ-TOMÁS, J.; CASELLES, V. Aplicação de redes neurais artificiais e regressão logística à previsão de incêndios florestais perigo na Galiza usando dados MODIS. **International Journal of Wildland Fire**, v. 21, n. 8, p. 1025-1029, 2012. DOI: 10.1071/WF11105.

BOWMAN, D. M. et al. A dimensão humana dos regimes de fogo na Terra. **International Journal of Wildland Fire**, v. 28, n. 7, p. 634-650, 2019.

BOWMAN, D. M. J. S.; MOREIRA-MUÑOZ, A.; KOLDEN, C. A.; JOHNSTON, F. H.; WILLIAMSON, G. J.; ROCHA, A. **Human–environmental drivers and impacts of the globally extreme 2017 Chilean fires**. Ambio, v. 48, n. 4, p. 352-362, 2019.

BRAGA, A. P.; CARVALHO, A. P. L. F.; LUDELMIR, T. B. **Redes Neurais Artificiais: Teoria e Aplicações**. 2. ed. Rio de Janeiro: LTC, 2014.

BRASIL. Ministério do Meio Ambiente. Conselho Nacional de Meio Ambiente. Resolução CONAMA nº 3, de 28 de junho de 1990. Diário Oficial da União, Brasília, DF, 22 ago. 1990.

BUI, D. T.; BUI, Q.-T.; NGUYEN, Q.-P.; PRADHAN, B.; NAMPAK, H.; TRINH, P. T. **A hybrid artificial intelligence approach using GIS-based neural-fuzzy inference system and particle swarm optimization for forest fire susceptibility modeling at a tropical area**. Agricultural and Forest Meteorology, v. 233, p. 32-44, 2017.

CARMO, C. N.; HACON, S. S. **Estudos de séries temporais de poluição atmosférica por queimadas e saúde humana**. Ciência & Saúde Coletiva, v. 18, n. 11, p. 3245-3258, 2013.

CASCIO, Wayne E. **Wildfire smoke and cardiovascular health**. Environmental Research, v. 241, p. 117578, 2024.

CASSOU, E. **Field Burning. Agricultural Pollution**. 2018. Disponível em: <https://openknowledge.worldbank.org/handle/10986/29504>. Acesso em: 20 ago. 2023.

CHANG, C. C.; LEE, T. Y.; WU, C. C. **A review of machine learning algorithms for fire management**. Expert Systems with Applications, v. 37, n. 12, p. 7591-7599, 2010.

CORPO DE BOMBEIROS MILITAR DE SANTA CATARINA. Tópicos introdutórios: ciências do fogo. Organizado por Marcos Alves da Silva. Florianópolis: Corpo de Bombeiros Militar de Santa Catarina, 2018.

DUARTE NETO, L. P.; NAKAYAMA, L. Y.; BAZZO, J. J. **Redes neurais para identificação de padrões na detecção de queimadas na floresta amazônica**. In: SIMPÓSIO BRASILEIRO DE COMPUTAÇÃO APLICADA, 1., 2011, Manaus. Anais [...]. Manaus: INPA, 2011.

EBBESEN, Lucas. Clima do Ceará. InfoEscola, [2016]. Disponível em: <https://www.infoescola.com/geografia/clima-do-ceara/>. Acesso em: 21 jul. 2023.

FARIAS, G. M.; SOUZA FILHO, F. A.; CASTRO, M. A. H.; SOUZA, D. L.; COSTA, L. H. M. **Aplicação de técnicas de machine learning no**

preenchimento de falhas em séries temporais de precipitação mensal. Revista DAE, São Paulo, v. 71, n. 242, p. 93-104, 2023.

FERNANDES, L. C. **Modelagem de risco de incêndios florestais utilizando redes neurais artificiais aplicada às regiões metropolitanas.** 2019. 163 f. Dissertação (Mestrado) – [Nome do Programa], [Nome da Universidade], [Local], 2019.

FERNEDA, E. Redes neurais e sua aplicação em sistemas de recuperação de informação. **Ciência da Informação**, Brasília, v. 32, n. 2, p. 45-53, 2003.

FOOD AND AGRICULTURE ORGANIZATION OF THE UNITED NATIONS (FAO). Global Forest Resources Assessment: Progress towards sustainable forest management. Roma: **FAO**, 2006. Disponível em: <http://www.fao.org/3/a-a0400e.pdf>. Acesso em: 10 jun. 2020.

FRANÇA, M. M.; FERNANDES FILHO, E. I.; XAVIER, B. T. L. Análise do uso da terra no município de Viçosa-MG mediado por classificações supervisionadas com Redes Neurais Artificiais e MAWVER. **Revista Brasileira de Geografia Física**, v. 2, p. 92-101, 2009.

GHORBANZADEH, O.; BLASCHKE, T.; GHOLAMNIA, K.; ARYAL, J. **Forest Fire Susceptibility and Risk Mapping Using Social/Infrastructural Vulnerability and Environmental Variables.** Fire, v. 2, n. 3, p. 50, 2019.

GIANNAKIDOU, S.; RADOGLOU-GRAMMATIKIS, P.; LAGKAS, T.; et al. Harnessing the power of the Internet of Things and artificial intelligence in wildfire prevention, detection, and restoration: A comprehensive survey. **Internet of Things**, v. 26, p. 101171, 2024.

GOMES, L. F. C.; MONTENEGRO, S. M. G. L.; VALENÇA, M. J. S. Modelo baseado na técnica de redes neurais para previsão de vazões na Bacia do Rio São Francisco. **Revista Brasileira de Recursos Hídricos**, v. 15, p. 5-15, 2010.

GONZALEZ, R. C.; WOODS, R. E. **Digital Image Processing.** 4. ed. Pearson, 2018.

GONZALEZ, Rafael C.; WOODS, Richard E. **Processamento Digital de Imagens.** 3. ed. São Paulo: Pearson Education, 2010.

HANG, H. T.; MALLICK, J.; ALQADHI, S.; BINDAJAM, A. A.; ABDO, H. G. Exploring forest fire susceptibility and management strategies in Western Himalaya: Integrating ensemble machine learning and explainable AI for accurate prediction and comprehensive analysis. **Environmental Technology & Innovation**, v. 35, p. 103655, 2024.

HAYKIN, S. **Redes neurais: princípios e prática.** Porto Alegre: Bookman, 2001.

INSTITUTO NACIONAL DE METEOROLOGIA (INMET). BDMep - Dados históricos. Brasília, DF: **INMET**, [20--]. Disponível em: <https://portal.inmet.gov.br/servicos/bdmep-dados-hist%C3%B3ricos>. Acesso em: 6 ago. 2024.

INSTITUTO NACIONAL DE PESQUISAS ESPACIAIS (INPE). BD queimadas. Brasília, DF: **INPE**, [20--]. Disponível em: <http://www.inpe.br/faq/index.php?pai=7>. Acesso em: 2 out. 2023.

JAIN, P.; COOGAN, S. C. P.; SUBRAMANIAN, S. G.; CROWLEY, M.; TAYLOR, S.; FLANNIGAN, M. D. A review of machine learning applications in wildfire science and management. **Environmental Reviews**, v. 28, p. 478-505, 2020.

JESUS, J. B. D.; ROSA, C. N. D.; BARRETO, Í. D. D. C.; FERNANDES, M. M. Análise da incidência temporal, espacial e de tendência de fogo nos biomas e unidades de conservação do Brasil. **Ciência Florestal**, Santa Maria, v. 30, n. 1, p. 176-191, 2020. DOI: 10.5902/1980509837696.

LEMES, G. P.; MATRICARDI, E. A. T.; COSTA, O. B.; LEAL, F. A. Avaliação espaço-temporal dos incêndios florestais no Parque Nacional Serra da Canastra no período de 1991 a 2011. **Ambiência**, Guarapuava, v. 10, p. 247-266, ago. 2014.

LIU, L.; CHEN, L.; ASADI, M. Capsule neural network and adapted golden search optimizer based forest fire and smoke detection. **Scientific Reports**, v. 15, n. 4187, p. 1-20, 2025.

LUZ, M. N. da; SOUTO, P. C.; HENRIQUE, G. da S.; DELFINO, R. de C. H.; LEITE, A. P. Comportamento do fogo em espécies nativas da Caatinga na região geográfica imediata de Patos-PB. **Ciência Florestal**, Santa Maria, v. 33, n. 3, e73573, p. 1-22, 2023.

MAEDA, E. E.; FORMAGGIO, A. R.; SHIMABUKURO, Y. E.; ARCOVERDE, G. F. B.; HANSEN, M. C. **Predicting forest fire in the Brazilian Amazon using MODIS imagery and artificial neural networks. International Journal of Applied Earth Observation and Geoinformation**, v. 11, p. 265-272, 2009.

MANICA, R. Aplicação de uma rede neural artificial simplificada para a identificação de gradação de depósitos turbidítico. **Geociências**, v. 32, p. 429-440, 2013.

MCCULLOCH, W. S.; PITTS, W. A logical calculus of the ideas immanent in nervous activity. **Bulletin of Mathematical Biology**, v. 5, p. 115–133, 1943.

MELO, A. V.; CIRNE, M. V. **Análise do programa de Brigadas em Pernambuco no contexto da importância das ações do manejo integrado do fogo**. Biodiversidade Brasileira, Brasília, n. 1, p. 299, 2019.

- MILLÁN, M. M. Extreme hydrometeorological events and climate change predictions in Europe. **Journal of Hydrology**, v. 518, p. 206-224, 2014.
- MORETTIN, P. A.; TOLOI, C. M. C. **Modelos para previsão de séries temporais**. Rio de Janeiro: IMPA, 1981.
- MORITZ, M. A.; PARISIEN, M. A.; BATLLORI, E.; KRAWCHUK, M. A.; HAYHOE, K.; SYPHARD, A. D.; HANDMER, J.; FINNEY, M. A. Climate change and disruptions to global fire activity. **Ecosphere**, v. 3, n. 6, p. 1-22, 2012.
- NASCIMENTO, R. L. X. N. et al. (org.). **Caderno de caracterização: estado do Ceará**. Brasília, DF: Codevasf, 2022.
- NATIONAL INTERAGENCY FIRE CENTER (NIFC). **Estatísticas de gestão de incêndios florestais**. 2020.
- NEVES, C. F. D. J. **Transição de incêndios florestais para mega-incêndios. Vulnerabilidades explicativas para o caso de Portugal continental**. 2020. Tese (Doutorado em Geografia) – Universidade de Coimbra, Coimbra, 2020.
- OLIVEIRA, R. C. de. **Redes Neurais Artificiais na solução de problemas inversos em engenharia – Estudo de Caso**. Iguatu CE: Quipá Editora 2024
- PARENTE, Laerte G.; PARENTE, Liana L.; PINTO, Livia K. de S.; ALENCAR, Ane. Assessing burned area patterns and drivers in the Brazilian Caatinga from 2001 to 2018. **Remote Sensing**, v. 13, n. 1, p. 65, 2021.
- PARKS, S. A.; HOLSINGER, L. M.; PANUNTO, M. H.; JOLLY, W. M.; DOBROWSKI, S. Z.; DILLON, G. K. High-severity fire: evaluating its key drivers and mapping its probability across western US forests. **Environmental Research Letters**, v. 13, 044037, 2018. DOI: 10.1088/1748-9326/aab791.
- PEREIRA, E. J. A. L.; FERREIRA, P. J. S.; LIMA, S. C. Policy in Brazil (2016–2019) threaten conservation of the Amazon rainforest. **Environmental Science & Policy**, v. 100, p. 11-14, 2019.
- PIVELLO, V. R. et al. Understanding Brazil's catastrophic fires: Causes, consequences and policy needed to prevent future tragedies. **Perspectives in Ecology and Conservation**, v. 19, n. 3, p. 233-255, 2021. Disponível em: <https://www.sciencedirect.com/science/article/pii/S2530064421000560>. Acesso em: 6 out. 2023.
- RODRIGUES, João A.; RAJÃO, Marcos G.; RIBEIRO, Lerry F.; FREITAS, Gabriel de O.; SOARES-FILHO, Britaldo S. A global-scale assessment of fire susceptibility and its drivers. **Science of The Total Environment**, v. 906, p. 167571, 2024.
- SANTOS, P. S. M.; DUARTE, A. C.; ROCHA-SANTOS, T. Microplastics in soils: assessment, analytics and risks. **Environmental Chemistry**, 2018.

SATHISHKUMAR, V. E.; CHO, J.; SUBRAMANIAN, M.; NAREN, O. S. **Forest fire and smoke detection using deep learning-based learning without forgetting.** *Fire Ecology*, v. 19, n. 9, p. 1-17, 2023.

SEYFFARTH, J. A. S.; RODRIGUES, V. **Impactos da seca sobre a biodiversidade da Caatinga.** *Parcerias Estratégicas*, v. 22, n. 44, p. 41-62, 2017.

SGORLA, H.; DE QUADROS, F. L. F.; CASSOL, H. L. G.; DE DAVID, C.; DOS SANTOS, J. S.; DA SILVA, A. C. **Modeling Fire Occurrence in the Pampa Biome**, Brazil. *Fire*, v. 7, n. 12, p. 437, 2024.

SILVA, A. C.; JUVANHOL, R. S.; MIRANDA, J. R. Variabilidade espaço-temporal de ocorrência e recorrência de fogo no Bioma Caatinga usando dados do sensor MODIS. **Ciência Florestal**, Santa Maria, v. 33, n. 1, e70195, p. 1-23, 2023. DOI: 10.5902/1980509870195.

SILVA, E. C. G.; FIEDLER, N. C.; NEVES, F. P.; CANZIAN, W. P. Gastos públicos com veículos e aeronaves empregados no combate ao incêndio florestal ocorrido na Reserva Biológica de Sooretama. **Ciência Florestal**, Santa Maria, v. 32, n. 3, p. 1290-1308, 2022. DOI: 10.5902/1980509861333.

SISTANI, T.; KAZEMITABAR, S.-J. Deep Learning Approach for Forest Fire Detection: A CNN Classification Model on the DeepFire Dataset. **Iranian Journal of Electrical and Electronic Engineering**, v. 21, n. 4, p. 3601, 2025.

SOARES, M. I.; RESCK, Z. M. R.; TERRA, F. S.; CAMELO, S. H. H. **Sistematização da assistência de enfermagem: facilidades e desafios do enfermeiro na gerência da assistência.** *Escola Anna Nery*, v. 19, n. 3, p. 491-498, 2015.

SOUZA, M. P.; COUTINHO, J. M. D. C. P.; SILVA, L. S.; AMORIM, F. S.; ALVES, A. R. Composição e estrutura da vegetação de caatinga no sul do Piauí, Brasil. **Revista Verde de Agroecologia e Desenvolvimento Sustentável**, v. 12, n. 2, p. 210-217, 2017.

VIGANÓ, R.; ZIMMERMANN, F. K.; PEREIRA, R. S. **Previsão de focos de incêndio no Pantanal utilizando Redes Neurais Artificiais.** In: SIMPÓSIO BRASILEIRO DE SENSORIAMENTO REMOTO, 18., 2017, Santos. **Anais** [...]. Santos: INPE, 2017. p. 3505-3512.

WU, C. Historical and future global burned area with changing. **ScienceDirect**, v. 2590332221001299, 2021.

ZIMMERMAN, J. B.; ANASTAS, P. T. Design through the 12 principles of green engineering. **Environmental Science & Technology**, v. 37, n. 5, p. 94A-101A, 2003

APÊNDICE A - CÓDIGO DO ALGORITMO DA REDE NEURAL FINAL

```

import pandas as pd
import numpy as np
from sklearn.model_selection import train_test_split
from sklearn.preprocessing import StandardScaler
from sklearn.impute import SimpleImputer
from sklearn.metrics import confusion_matrix, classification_report,
accuracy_score, precision_score
import tensorflow as tf
from tensorflow.keras.models import Sequential
from tensorflow.keras.layers import Dense
import matplotlib.pyplot as plt
import seaborn as sns
import os

# Criar diretório para salvar gráficos
if not os.path.exists("graphs_reduced_features"):
    os.makedirs("graphs_reduced_features")

# Carregar os dados
df = pd.read_csv("DadosFelipeFinal.csv", parse_dates=["DATA"])

# Criar a variável MES_CRITICO (Julho, Agosto, Setembro, Outubro)
dff["MES_CRITICO"] = df["DATA"].dt.month.isin([7, 8, 9, 10]).astype(int)

# Selecionar as 5 variáveis meteorológicas e a nova variável MES_CRITICO
selected_features = ["RADGLB", "UMIDREL", "TEMPMAX", "PRECIP", "VELVENT",
"MES_CRITICO", "FOGO"]
df_filtered = df[selected_features].copy()

# Verificar valores nulos e preencher com média
imputer = SimpleImputer(strategy="mean")
df_imputed = pd.DataFrame(imputer.fit_transform(df_filtered.drop("FOGO", axis=1)),
columns=df_filtered.drop("FOGO", axis=1).columns)

# Separar features e target
X = df_imputed
y = df_filtered["FOGO"]

# Divisão dos dados: 70% teste, 15% treino, 15% validação X_temp, X_test,
y_temp, y_test = train_test_split(X, y, test_size=0.3, random_state=42)
X_train, X_val, y_train, y_val = train_test_split(X_temp, y_temp,
test_size=0.5,
random_state=42)

# Padronização dos dados
scaler = StandardScaler()
X_train_scaled = scaler.fit_transform(X_train)
X_val_scaled = scaler.transform(X_val)
X_test_scaled = scaler.transform(X_test)

# Parâmetros para testar
activations = ["relu", "sigmoid", "tanh"]
neurons_list = [5, 10, 15, 20]
results = []
# Loop sobre diferentes configurações
for activation in activations:
    for neurons in neurons_list:

```

```

model = Sequential([
    Dense(neurons, activation=activation, input_shape=
(X_train.shape[1],)),
    Dense(1, activation="sigmoid")
])
model.compile(optimizer="adam", loss="binary_crossentropy", metrics=
["accuracy"])

# Treinar o modelo
history = model.fit(X_train_scaled, y_train,
                     validation_data=(X_val_scaled, y_val),
                     epochs=50, verbose=0, batch_size=32)

# Avaliar no conjunto de validação
val_loss, val_acc = model.evaluate(X_val_scaled, y_val, verbose=0)
y_val_pred = (model.predict(X_val_scaled) > 0.5).astype(int)
val_precision = precision_score(y_val, y_val_pred)

# Avaliar no conjunto de treino para verificar overfitting
train_acc = history.history["accuracy"][-1]
results.append({
    "activation": activation,
    "neurons": neurons,
    "val_accuracy": val_acc,
    "val_precision": val_precision,
    "train_accuracy": train_acc,
    "history": history,
    "model": model
})

# Gerar gráficos de acurácia e loss para cada modelo testado
plt.figure(figsize=(10, 4))
plt.subplot(1, 2, 1)
plt.plot(history.history["accuracy"], label="Treino")
plt.plot(history.history["val_accuracy"], label="Validação")
plt.title(f"Acurácia ({activation}, {neurons} neurônios)")
plt.legend()

plt.subplot(1, 2, 2)
plt.plot(history.history["loss"], label="Treino")
plt.plot(history.history["val_loss"], label="Validação")
plt.title(f"Loss ({activation}, {neurons} neurônios)")
plt.legend()
plt.tight_layout()

plt.savefig(f"graphs_reduced_features/performance_{activation}_{neurons}_neurons.png")
plt.close()

# Encontrar melhor modelo com base na acurácia de validação
best_result = max(results, key=lambda x: x["val_accuracy"])
print(f"Melhor modelo: {best_result['activation']} com {best_result['neurons']} neurônios")
print(f"Acurácia validação: {best_result['val_accuracy']:.4f}")
print(f"Precisão validação: {best_result['val_precision']:.4f}")

# Treinar melhor modelo com todos os dados de treino+validação
X_final_train = np.vstack([X_train_scaled, X_val_scaled])
y_final_train = np.hstack([y_train, y_val])
best_model = Sequential([
Dense(best_result["neurons"], activation=best_result["activation"]),
    input_shape=(X_train.shape[1],)),
    Dense(1, activation="sigmoid")
])

```

```

best_model.compile(optimizer="adam", loss="binary_crossentropy", metrics=
["accuracy"])
best_model.fit(X_final_train, y_final_train, epochs=50, verbose=0,
batch_size=32)

# Avaliar no conjunto de teste
test_loss, test_acc = best_model.evaluate(X_test_scaled, y_test, verbose=0)
y_test_pred = (best_model.predict(X_test_scaled) > 0.5).astype(int)
test_precision = precision_score(y_test, y_test_pred)
print(f"\nDesempenho no conjunto de teste:")
print(f"Acurácia: {test_acc:.4f}")
print(f"Precisão: {test_precision:.4f}")

# Matriz de confusão para validação
y_val_pred_best = (best_result["model"].predict(X_val_scaled) >
0.5).astype(int)
cm_val = confusion_matrix(y_val, y_val_pred_best)
plt.figure(figsize=(8, 6))
sns.heatmap(cm_val, annot=True, fmt="d", cmap="Blues")
plt.title("Matriz de Confusão - Validação")
plt.ylabel("Verdadeiro")
plt.xlabel("Preditivo")
plt.savefig("graphs_reduced_features/confusion_matrix_validation.png") # Salvar a imagem
plt.close()

# Matriz de confusão para teste
cm_test = confusion_matrix(y_test, y_test_pred)
plt.figure(figsize=(8, 6))
sns.heatmap(cm_test, annot=True, fmt="d", cmap="Blues")
plt.title("Matriz de Confusão - Teste")
plt.ylabel("Verdadeiro")
plt.xlabel("Preditivo")
plt.tight_layout()
plt.savefig("graphs_reduced_features/confusion_matrix_test.png") # Salvar a imagem
plt.close()

# Gráficos de desempenho do melhor modelo
plt.figure(figsize=(12, 4))
plt.subplot(1, 2, 1)
plt.plot(best_result["history"].history["accuracy"], label="Treino")
plt.plot(best_result["history"].history["val_accuracy"], label="Validação")
plt.title("Acurácia durante o Treinamento")
plt.legend()

plt.subplot(1, 2, 2)
plt.plot(best_result["history"].history["loss"], label="Treino")
plt.plot(best_result["history"].history["val_loss"], label="Validação")
plt.title("Loss durante o Treinamento")
plt.legend()
plt.tight_layout()
plt.savefig("graphs_reduced_features/performance_best_model.png") # Salvar a imagem do melhor modelo
plt.close()

# Verificar overfitting/underfitting
if best_result["train_accuracy"] > best_result["val_accuracy"] + 0.1:
    print("Possível overfitting detectado")
elif best_result["val_accuracy"] < 0.6:
    print("Possível underfitting detectado")
else:
    print("Modelo com bom equilíbrio entre treino e validação")

```

```

# Relatório de classificação detalhado
print("\nRelatório de Classificação (Teste):")
classification_report_str = classification_report(y_test, y_test_pred,
output_dict=True)
print(classification_report(y_test, y_test_pred))

#####
##Importância das Variáveis

# Criar e treinar o modelo para importância das variáveis (usando a mesma configuração do melhor modelo)
model_importance = Sequential([
    Dense(best_result["neurons"], activation=best_result["activation"]),
    input_shape=(X_train.shape[1],),
    Dense(1, activation="sigmoid")
])
model_importance.compile(optimizer="adam", loss="binary_crossentropy", metrics=[
    "accuracy"])
model_importance.fit(X_train_scaled, y_train, validation_data=(X_val_scaled, y_val),
epochs=50, verbose=0, batch_size=32)

# Calcular a importância das variáveis
weights = model_importance.layers[0].get_weights()[0] # Pesos da primeira camada
importance = np.mean(np.abs(weights), axis=1) # Média absoluta dos pesos por feature

# Normalizar a importância para percentuais
importance = 100 * importance / np.sum(importance)

# Criar DataFrame com as importâncias
feature_importance = pd.DataFrame({
    "Feature": X.columns,
    "Importance": importance
}).sort_values("Importance", ascending=True)

# Plotar o gráfico de importância
plt.figure(figsize=(10, 6))
plt.barh(feature_importance["Feature"], feature_importance["Importance"],
color="skyblue")
plt.xlabel("Importância Relativa (%)")
plt.title("Importância das Variáveis no Modelo de Previsão de Incêndios (Features Reduzidas)")
plt.grid(axis="x", alpha=0.3)
# Adicionar valores percentuais nas barras
for i, v in enumerate(feature_importance["Importance"]):
    plt.text(v + 0.5, i, f"{v:.1f}%", va="center")
plt.tight_layout()
plt.savefig("graphs_reduced_features/feature_importance_plot_reduced_features.png # Salvar a imagem")
plt.close()

# Exibir a importância em forma de tabela
print("\nImportância das Variáveis (Features Reduzidas):")
print(feature_importance.sort_values("Importance", ascending=False))

# Gráficos de dispersão para TODAS as variáveis importantes
all_important_features = feature_importance["Feature"].tolist()

for feature in all_important_features:
    plt.figure(figsize=(10, 6))
    sns.scatterplot(x=df_imputed[feature], y=y, hue=y, palette="viridis", alpha=0.6)

```

```
plt.title(f"Dispersão de {feature} vs. Ocorrência de Fogo (Features Reduzidas)")
plt.xlabel(feature)
plt.ylabel("FOGO (0=Não Fogo, 1=Fogo)")
plt.yticks([0, 1], ["Não Fogo", "Fogo"])

plt.savefig(f"graphs_reduced_features/scatter_{feature}_vs_FOGO_reduced_features. plt.close()

# Salvar o relatório de classificação em um arquivo para fácil acesso with
open("graphs_reduced_features/classification_report_test_reduced_features.txt", "w") as f:
    f.write(classification_report(y_test, y_test_pred))
```