

BRUNO NOLETO DE SOUSA

UTILIZAÇÃO DE REDES NEURAIS ARTIFICIAIS PARA A CLASSIFICAÇÃO DE RISCO DE INCÊNDIOS FLORESTAIS







BRUNO NOLETO DE SOUSA

UTILIZAÇÃO DE REDES NEURAIS ARTIFICIAIS PARA A CLASSIFICAÇÃO DE RISCO DE INCÊNDIOS FLORESTAIS

Trabalho de Conclusão de Curso apresentado ao curso de Bacharelado em Ciência da Computação da Universidade Estadual do Paraná para obtenção do título de Bacharel em Ciência da Computação.

Orientador: Prof. Me. Fabio Takeshi Matsunaga

SOUSA, B. N. de

Utilização de Redes Neurais Artificiais Para a Classificação de Risco de Incêndios Florestais/ Bruno Noleto de Sousa. – Apucarana, 2019.

57 p.: il. (algumas color.).

Trabalho de Conclusão de Curso – Bacharelado em Ciência da Computação – Universidade Estadual do Paraná, 2019.

Orientador: Prof. Me. Fabio Takeshi Matsunaga

1. Queimadas 2. Classificação de Risco 3. Aprendizado de Máquina. 4. Redes Neurais Artificiais. 4. Perceptron Multicamadas. I. Prof. Me. Fabio Takeshi Matsunaga. II. Universidade Estadual do Paraná. III. Centro de Ciências Sociais Aplicadas. IV. Título

CDU 004.8

BRUNO NOLETO DE SOUSA

UTILIZAÇÃO DE REDES NEURAIS ARTIFICIAIS PARA A CLASSIFICAÇÃO DE RISCO DE INCÊNDIOS FLORESTAIS

Trabalho de Conclusão de Curso apresentado ao curso de Bacharelado em Ciência da Computação da Universidade Estadual do Paraná para obtenção do título de Bacharel em Ciência da Computação.

BANCA EXAMINADORA

Prof. Me. Fabio Takeshi Matsunaga Universidade Estadual do Paraná Orientador

Prof. Me. Edison Antonio Sahd Filho Universidade Estadual do Paraná

Prof. Me. Marcelo Caetano de Cernev Rosa Universidade Estadual do Paraná

Apucarana, 27 de novembro de 2019

icado a todos os que		n nas pessoas, incentiva	ındo
independente de se	u nível intelectual oi	social.	
independente de se	u nível intelectual oi	ı social.	
independente de se	u nível intelectual oi	social.	
independente de se	u nível intelectual oi	social.	
independente de se	u nível intelectual oi	social.	

AGRADECIMENTOS

Agradeço este trabalho à minha família, principalmente aos meus pais, que sempre estiveram ao meu lado dando suporte às difíceis decisões que foram necessárias a partir do momento que fui aprovado em um vestibular a milhares de quilômetros de casa. E principalmente por nunca deixarem de acreditar em mim mesmo em momentos mais turbulentos.

Agradeço também à todos os professores que são apaixonados por sua profissão. Pessoas que acreditaram não apenas em mim, mas também em todos os seus alunos de graduação. Inspirado por estas personalidades incríveis, fui capaz de melhorar como pessoa, de ter forças para não desistir no meio do caminho, e também fui capaz de descobrir a trilha para o conhecimento, algo necessário para realizar trabalhos como esta dissertação.

Também gostaria de agradecer especialmente ao Vinícius. Por sempre me dar todo o apoio emocional e psicológico, estando comigo nos momentos mais difíceis, mesmo morando em outra cidade.

Agradeço ao meu orientador Fabio Matsunaga, por confiar no desenvolvimento do meu trabalho e estar sempre à disposição para solucionar questões e dúvidas.

Sou grato a todos os colegas de diferentes turmas, tanto os que já obtiveram a graduação quanto os que continuam suas nobres jornadas dentro ou fora da UNESPAR.

Agradeço aos meus amigos mais antigos que estão a milhares de quilômetros de distância Matheus Matalobos, Sophia Santos, Thiago Cerqueira, Vinícius Batitucci e Vitor Santos, e aos muitos outros amigos nos quais eu fiz nesta caminhada.

E por fim, mas não menos importante, agradeço e dedico este trabalho especialmente aos amigos de *cosplay* Carina Silva (*in memoriam*) e Kleder Zamberlan (*in memoriam*).

Sem o apoio de todas estas pessoas, minha caminhada até este momento teria sido impossível.

SOUSA, B. N. DE. Utilização de Redes Neurais Artificiais Para a Classificação de Risco de Incêndios Florestais. 57 p. Trabalho de Conclusão de Curso (Bacharelado em Ciência da Computação) – Universidade Estadual do Paraná, Apucarana, 2019.

RESUMO

A maioria das ações do poder público em relação à prevenção de incêndios florestais são baseadas no cálculo do Índice de Risco de Fogo do Instituto Nacional de Pesquisas Espaciais. Este trabalho propõe um método alternativo de classificação de risco de incêndio florestal, utilizando conceitos de aprendizado de máquina, dentro da qual será realizado o treinamento de uma rede neural artificial. Utilizando dois vastos bancos de dados meteorológicos (BDMEP) e de queimadas (BDQueimadas), serão filtradas e utilizadas algumas de suas variáveis para ao final determinar o índice de risco de incêndio, dada uma condição meteorológica de um município específico. A rede neural ao final de seu treinamento obteve 93,99% de acurácia e 87,19% de precisão para as amostras de validação. Ao analisar a taxa de acertos para incêndios registrados, o modelo da rede neural obteve 70,32% de acertos.

Palavras-chave: Queimadas. Classificação de Risco. Aprendizado de Máquina. Redes Neurais Artificiais. Perceptron Multicamadas.

SOUSA, B. N. DE. **Use of Artificial Neural Networks for Wildfire Risk Classification**. 57 p. Final Project (Bachelor of Science in Computer Science) – State University of Paraná, Apucarana, 2019.

ABSTRACT

Most of the actions of the public authorities regarding forest fire prevention are based on the calculation of the Fire Risk Index of the National Institute for Space Research. This work proposes an alternative method of forest fire risk classification, using machine learning concepts, in which the training of an artificial neural network will be performed. Using two vast weather (BDMEP) and wildfire (BDQueimadas) databases, some of its variables will be filtered and used to determine the fire risk index, given a weather condition of a specific municipality. The neural network at the end of its training obtained 93.99% accuracy and 87.19% precision for the validation set. By analyzing the hit rate for recorded fires, the neural network model obtained 70.32% hits.

Keywords: Wildfires. Risk Classification. Machine Learning. Artificial Neural Networks. Multilayer Perceptron.

LISTA DE ILUSTRAÇÕES

Figura 1 – Modelo de um neurônio k	25
Figura 2 – Modelo de uma <i>Perceptron</i> de camadas múltiplas (2 camadas ocultas)	26
Figura 3 – Representação visual da função de perda	27
Figura 4 – Diagrama da utilização das duas bases de dados para gerar o <i>dataset</i>	33
Figura 5 – Exemplo da preparação e resultado da junção dos dados	36
Figura 6 – Localização de Imperatriz/MA na "Amazônia Legal"	41
Figura 7 – Gráfico da função de perda durante o treinamento	43
Figura 8 – Gráfico da acurácia durante o treinamento	44
Figura 9 – Gráfico da precisão durante o treinamento	44
Figura 10 – Gráfico da precisão para a classe 4 (MÁXIMO)	45
Figura 11 – Gráficos com a indicação da presença de fogo referente ao agrupamento de	
classes calculadas pelo INPE	46
Figura 12 – Gráficos com a indicação da presença de fogo referente ao agrupamento de	
classes calculadas pela MLP	46
Figura 13 – Quantidade de amostras para cada classe quando houve fogo	47
Figura 14 – Quantidade de amostras para cada classe quando não houve fogo	47

LISTA DE TABELAS

Tabela 1 – Queimadas Detectadas no Brasil via Satélite pelo INPE	23
Tabela 2 – Classes de Risco de Fogo a partir do RF calculado	29
Гаbela 3 — Valores assumidos pela constante "A" para as 7 classes de vegetação do Risco	
de Fogo	29
Tabela 4 – Definição das variáveis registradas em cada medição	34
Tabela 5 – Verificação INPE	48
Tabela 6 – Verificação MLP	48

LISTA DE ABREVIATURAS E SIGLAS

AMD Advanced Micro Devices, Inc.

BDMEP Banco de Dados Meteorológicos para Ensino e Pesquisa

BDQueimadas Banco de Dados de Queimadas

DL4J Deep Learning for Java

GB Gigabyte

IBGE Instituto Brasileiro de Geografia e Estatística

INMET Instituto Nacional de Meteorologia

INPE Instituto Nacional de Pesquisas Espaciais

MLP Perceptron de Múltiplas Camadas (*Multilayer Perceptron*)

RAM Memória de Acesso Aleatório (*Random Access Memory*)

RF Índice de Risco de Fogo

RNA Rede Neural Artificial

UTC Tempo Coordenado Universal (*Universal Time Coordinated*)

SUMÁRIO

1	INTRODUÇAO	23
2	FUNDAMENTAÇÃO TEÓRICA	25
2.1		25
2.1.1	Perceptrons	25
2.2	Técnicas de análise do desempenho da RNA	26
2.2.1	Função de perda (LOSS)	26
2.2.2	Validação <i>hold-out</i>	27
2.2.3	Acurácia	27
2.2.4	Precisão	27
2.3	Índice de Risco de Fogo (INPE)	28
2.3.1		29
2.3.2		30
2.3.3	_	30
2.4		30
3	MÉTODO DE PESQUISA	33
3.1		33
3.2		35
3.2.1	Seleção/extração de características	37
3.2.2	Normalização do <i>dataset</i>	37
3.2.3	Separação dos <i>datasets</i> de treinamento e validação	37
3.2.4	Balanceamento das amostras de treinamento	37
3.3	Treinamento	38
3.4		40
3.5	_	40
4	EXPERIMENTO	41
4.1	Dados de entrada	42
4.2	Treinamento	42
5	RESULTADOS	43
5.1	Treinamento	43
5.2	Análise de acerto para incêndios registrados	45
6	CONCLUSÃO	51

REFERÊNCIAS	53
ANEXOS	55
ANEXO A – FÓRMULAS DOS FATORES DE PRECIPITAÇÃO	57

1 INTRODUÇÃO

Os incêndios florestais são um problema recorrente no mundo. Independente do agente causador, é sinônimo de problemas tanto para o meio ambiente, quanto para moradores de áreas atingidas e o próprio poder público. Como exemplo, temos o grande incêndio no estado da Califórnia (EUA), ocorrido em novembro de 2018. Apenas neste evento, cerca de 620 mil km² foram tomados pelo fogo, causando 85 fatalidades e fazendo com que 52 mil pessoas fossem evacuadas da região. Além de provocar prejuízos econômicos diretos e indiretos (CNBC, 2018)(MCBRIDE et al., 2018).

No Brasil, foram registrados pelo Instituto Nacional de Pesquisas Espaciais (INPE) 132.955 focos de incêndio florestal no ano de 2018. Embora a quantidade observada seja menor que a de 2017, a série histórica se mostra bastante variável. Na Tabela 1, são demonstradas as variações entre 2017 e 2018, e entre 2018 e 2019, utilizando os dados dos meses de Janeiro e Fevereiro.

Tabela 1 – Queimadas Detectadas no Brasil via Satélite pelo INPE

Ano	Mês			Total Anual
Ano -	Janeiro	Fevereiro	Jan+Fev	10tai Alluai
2017	2.255	1.245	3.500	207.633
2018	2.560	1.479	4.039	132.955
Variação	+13,53%	+18,80%	+15,40%	-35,97%

Ano	Mês			Total Anual	
Allo	Janeiro Fevere		Jan+Fev	· Iotal Allual	
2018	2.560	1.479	4.039	132.955	
2019	4.030	2.865	6.895	N/D	
Variação	+57,42%	+93,71%	+70,71%	N/D	

Fonte: Portal Queimadas do INPE

Apenas nos dois primeiros meses de 2019, foi observado um aumento de aproximadamente 70,71% de queimadas detectadas em comparação ao mesmo intervalo de 2018¹. Portanto, existe a necessidade de que estudos sejam realizados, para que seja possível aumentar a efetividade das políticas públicas em face deste problema. Segundo Fonseca-Morello et al. (2017), maior parte do orçamento anual do programa PREVFOGO é destinado ao combate de incêndios, ao invés da sua prevenção. Uma das razões apontadas se deve ao fato de ser mais fácil chamar a atenção das autoridades utilizando dados estatísticos plenamente observados, em detrimento de dados probabilísticos.

Dados do Portal Queimadas do Instituto Nacional de Pesquisas Espaciais. Disponível em: http://www.inpe.br/queimadas/portal/estatistica_paises. Acesso em: 05 mar. 2019

Porém, a constante evolução do poder de processamento computacional culminou na evolução da área da Inteligência Artificial. Consequentemente, abriu-se um novo leque de métodos para o uso das informações obtidas. Assim, a sociedade passou a ter mais outros meios de tratar problemas até então difíceis, ou até mesmo impossíveis de serem evitados.

Considerando a situação exposta, o objetivo deste trabalho é classificar o risco de incêndio florestal para uma localidade específica, a partir de informações existentes em bancos de dados públicos meteorológicos e de incêndios, utilizando conhecimentos da área da Inteligência Artificial, tal como modelos de aprendizado de máquina, dentro da qual podemos citar as Redes Neurais Artificiais (RNA).

Este trabalho está organizado da seguinte forma: no Capítulo 2 serão apresentados os conceitos que serão utilizados e trabalhos correlatos na área. No Capítulo 3 serão apresentados os métodos e técnicas que serão utilizados neste trabalho. No Capítulo 4 será abordado o experimento que foi realizado para este trabalho. No Capítulo 5, os dados de saída e a análise de acerto do experimento serão apresentados. E no Capítulo 6 será apresentada a conclusão com base nos resultados obtidos por este trabalho.

2 FUNDAMENTAÇÃO TEÓRICA

Para o entendimento integral deste trabalho, são necessários alguns conhecimentos básicos de conceitos da Inteligência Artificial. Além disso, também é necessária a exposição de estudos semelhantes feitos por outros autores, ou até mesmo métodos externos à computação que serão utilizados para comparação nos capítulos posteriores. Esta seção visa dar a base de conhecimento necessária para a compreensão dos métodos a serem utilizados.

2.1 Redes Neurais Artificiais

Segundo Haykin (2009), uma Rede Neural Artificial (RNA) é um processador de dados distribuído paralelamente, moldado a partir de pequenas e simples unidades de processamento. Foi concebida de maneira que pudesse reproduzir virtualmente o método de funcionamento do cérebro humano (HAYKIN, 2009). Da mesma maneira que temos os neurônios no sistema nervoso biológico, nas RNAs temos os núcleos de processamento, que aqui também serão chamados de neurônios, para que na etapa de seu treinamento seja possível cada núcleo "aprender" um determinado dado com base em um algoritmo especificado (HAYKIN, 2009). A Figura 1 demonstra um modelo de um neurônio artificial.

Sinais de entrada $\begin{array}{c} x_1 & & & \\ x_2 & & & \\ x_3 & & & \\ x_4 & & & \\$

Figura 1 – Modelo de um neurônio *k*

Fonte: Haykin (2009)

2.1.1 Perceptrons

A *Perceptron* foi a primeira RNA a ser descrita na forma de algoritmo (HAYKIN, 2009). Foi idealizada por Rosenblatt em meados do século XX (HAYKIN, 2009). Ela possuía uma única camada de processamento (HAYKIN, 2009). A partir da *Perceptron* de Rosenblatt, posteriormente foi concebida sua variante de múltiplas camadas. Uma das diferenças da *Perceptron*

simples para o seu modelo de multicamadas é que durante seu treinamento é utilizado um algoritmo de retropropagação de erro (HAYKIN, 2009). Além disso, diferentemente da *Perceptron* simples, na multicamadas é possível resolver problemas não-lineares, como o problema do "ou-exclusivo" (HAYKIN, 2009). A Figura 2 apresenta um modelo de RNA de múltiplas camadas.

Sinal de entrada

Camada camada camada oculta

Figura 2 – Modelo de uma *Perceptron* de camadas múltiplas (2 camadas ocultas)

Fonte: Haykin (2009)

2.2 Técnicas de análise do desempenho da RNA

Para obter o modelo de RNA mais adequado para a resolução do problema proposto, são utilizadas diversas métricas bastante conhecidas pela comunidade científica.

2.2.1 Função de perda (*LOSS*)

Utilizando um conceito semelhante à estatística, é escolhida uma função específica para calcular a diferença entre valores previstos e esperados (WIKIMEDIA, 2019). Na área de aprendizado de máquina, este mesmo entendimento é utilizado, para calcular a diferença entre a saída esperada e a saída obtida pela RNA (PARMAR, 2018).

Quando o valor de uma função de perda para uma amostra é igual a 0, significa que a rede acertou perfeitamente sua previsão com base na saída esperada. Caso contrário, significa que ela errou sua previsão (DEVELOPERS, 2019). As Figuras 3a e 3b demonstram graficamente o que seria o resultado das funções de perda de dois modelos hipotéticos. A linha tracejada de cada figura é a saída obtida, e o comprimento das setas são exatamente a função de perda para cada amostra (DEVELOPERS, 2019).

y (a) Função de perda alta (b) Função de perda baixa

y X1 X1

Figura 3 – Representação visual da função de perda

Fonte: Google Developers (2019)

2.2.2 Validação hold-out

Estimar o quanto uma RNA está acertando em suas amostras é importante tanto para verificar se ela possui tendência a melhorar ou piorar, quanto para também escolher o "melhor" modelo possível dentre todos os modelos de rede testados (WOLPERT, 1992)(KOHAVI et al., 1995).

O método de validação *hold-out*, também chamada de validação cruzada por retenção, separa o *dataset* em dois conjuntos mutuamente exclusivos (RUSSELL; NORVIG, 2010). Um conjunto será utilizado para o treinamento da rede, e o outro apenas para a validação (YADAV; SHUKLA, 2016). Durante a etapa de treinamento, as amostras do conjunto de treinamento são utilizadas para ajustar o modelo da RNA, e as amostras pertencentes ao conjunto de validação são testadas no modelo corrente da RNA para verificar a capacidade de generalização da rede (HAYKIN, 2009)(YADAV; SHUKLA, 2016).

2.2.3 Acurácia

A acurácia consiste no cálculo da taxa resultante da soma de todos os valores previstos corretamente em relação a todo o *dataset* (PROVOST; KOHAVI, 1998). O cálculo da acurácia é demonstrado na Equação 2.1.

$$Acurácia = \frac{Qtd. \ Acertos}{Qtd. \ Total \ de \ Amostras \ do \ \textit{Dataset}}$$
 (2.1)

2.2.4 Precisão

A precisão indica a proporção de valores classificados corretamente dentre todos os valores obtidos como uma classe determinada (PROVOST; KOHAVI, 1998). A fórmula para

calcular a taxa de precisão é apresentada na Equação 2.2.

$$Precisão = \frac{Qtd. Acertos da Classe}{Qtd. Total de Amostras da Classe}$$
(2.2)

A precisão média para todo o modelo é calculada a partir da média aritmética simples da precisão de todas as classes existentes. Sua fórmula é demonstrada na Equação 2.3. Considerando que n é a quantidade de classes existentes no modelo, i é o índice da classe atual e x_i é o valor da precisão da classe de índice i.

Precisão Média =
$$\frac{1}{n} \cdot \sum_{i=1}^{n} x_i$$
 (2.3)

2.3 Índice de Risco de Fogo (INPE)

O índice de Risco de Fogo (RF) foi desenvolvido pelo Centro de Previsão de Tempo e Estudos Climáticos do Instituto Nacional de Pesquisas Espaciais (INPE), a partir de dados registrados de uma quantidade considerável de incêndios florestais nos principais biomas e tipos de vegetação no Brasil nas últimas décadas (SISMANOGLU; SETZER, 2004).

O cálculo do RF baseia-se em três variáveis meteorológicas principais, nas quais são utilizados o acumulado de precipitação, umidade relativa do ar mínima e temperatura máxima. Segundo Setzer et al. (2016), a velocidade do vento é desconsiderada do cálculo, por influenciar apenas na propagação do fogo, e não na sua ignição.

Para realizar o cálculo final, são realizados outros cálculos preliminares para que seja possível ponderar cada variável meteorológica. A equação 2.4 demonstra a fórmula para o RF observado. Em que RB é o "risco básico", FT é o fator de precipitação e FU é o fator de umidade. As subseções 2.3.1, 2.3.2 e 2.3.3 demonstram como são feitos os cálculos preliminares de cada variável desta equação.

$$RF = RB \cdot FT \cdot FU \tag{2.4}$$

A partir do resultado desta equação, o valor será atribuído em uma das 5 classes possíveis de risco, como pode ser visto na Tabela 2.

Tabela 2 – Classes de Risco de Fogo a partir do RF calculado

Classe	Valor
Mínimo	$RF \le 0,15$
Baixo	$0,15 < RF \le 0,4$
Médio	$0, 4 < RF \le 0, 7$
Alto	$0,7 < RF \le 0,95$
Crítico	RF > 0,95

Fonte: Setzer et al. (2016)

2.3.1 Risco Básico (RB)

A variável de Risco Básico nunca ultrapassará o valor de 0.9, e ela aumenta a partir de uma curva senoidal, pois a variação da intensidade e duração da luz do Sol durante o ano também possui um padrão senoidal (SETZER et al., 2016). A variável RB é calculada por meio da equação 2.5:

$$RB = \frac{0.9 \cdot \{1 + sen[((A \cdot PSE) - 90) \cdot \frac{3.14}{180}]\}}{2}$$
 (2.5)

Porém, para o cálculo utilizando esta fórmula, é necessário obter o valor de outras duas variáveis, que aqui são denominadas como a constante A e a variável PSE.

A constante A é definida pelo tipo de vegetação da localidade que será calculado o risco de fogo (SETZER et al., 2016). Os valores para cada classe de vegetação são demonstrados na Tabela 3.

Tabela 3 – Valores assumidos pela constante "A" para as 7 classes de vegetação do Risco de Fogo

Classe de Vegetação	Valor
Ombrófila Densa; Alagados	1,5
Florestas Decíduas e Sazonais	1,72
Florestas Contato; Campinarana	2,0
Savana Arbórea; Caatinga Fechada	2,4
Savana; Caatinga Aberta	3,0
Agricultura e Diversos	4,0
Pastagens Gramíneas	6,0

Fonte: Setzer et al. (2016)

A variável PSE é obtida a partir da ponderação dos acumulados de chuva em diferentes intervalos até os 120 dias anteriores ao dia da medição (SETZER et al., 2016). Como define Setzer et al. (2016), uma precipitação de alguns milímetros de um tempo mais recente possui efeito redutor na secura muito maior do que uma precipitação ocorrida muito tempo atrás.

Apesar da variável também ser chamada de "dias de secura", ela corresponde ao cálculo de uma quantidade **hipotética** de dias consecutivos sem chuva (SETZER et al., 2016). O cálculo do PSE é demonstrado na equação 2.6.

$$PSE = 105 \cdot fp1 \cdot fp2 \cdot fp3 \cdot fp4 \cdot fp5 \cdot fp6a10 \cdot fp11a15 \cdot fp16a30 \cdot fp31a60 \cdot fp61a90 \cdot fp91a120$$
(2.6)

Em que cada variável desta equação é referente à um "fator de precipitação" para um intervalo de tempo específico anterior ao dia a ser medido (SETZER et al., 2016). A fórmula para cada fator de precipitação é definida no Anexo A.

2.3.2 Fator de Temperatura (FT)

O cálculo do FT é realizado conforme a Equação 2.7.

$$FT = [T_{max} \cdot (0,02)] + 0,4 \tag{2.7}$$

Em que a variável T_{max} é definida pela temperatura máxima do dia de interesse (SETZER et al., 2016).

2.3.3 Fator de Umidade (FU)

O fator de umidade FU é calculado por meio da Equação 2.8.

$$FU = [U_{min} \cdot (-0,006)] + 1,3 \tag{2.8}$$

Em que a variável U_{min} é definida pela umidade relativa do ar mínima do dia em questão (SETZER et al., 2016).

2.4 Trabalhos correlatos

Diversos estudos com objetivos semelhantes foram desenvolvidos pela comunidade científica na área de Inteligência Artificial e/ou Matemática, com alguns pesquisadores utilizando dados meteorológicos, outros utilizando dados obtidos via satélite, e em alguns casos até ambos.

Sayad et al. (2019) realizaram experimentos utilizando *Big Data*, aprendizado de máquina, sensoriamento remoto e algoritmos de *Data Mining*. Cada saída indicava a presença ou não de fogo, ou seja, sua saída era binária. Seu trabalho obteve uma taxa de acerto alta em cada um dos dois algoritmos de aprendizado de máquina testados. Em comparação aos dados reais de incêndios florestais registrados na localização estudada, a acurácia foi 98,32% ao utilizar um modelo MLP, e de 97,48% ao utilizar uma Máquina de Vetores de Suporte (SVM).

Silva et al. (2018) elaboraram uma classificação de risco e propagação de incêndios florestais utilizando conceitos matemáticos. No seu trabalho foi utilizado um sistema dinâmico *Fuzzy* e um classificador do tipo *K-Nearest Neighbors* (k-NN) para o agrupamento dos resultados indicados. Sua taxa de acerto foi de 73,97%.

Lall e Mathibela (2016) realizaram um estudo sobre a utilização de redes neurais artificiais para a classificação de risco para a Cidade do Cabo, na África do Sul. Foram utilizados dados do ano de 2009 até o ano de 2015, levando em consideração variáveis climáticas, topológicas e de vegetação para classificar o risco de incêndio em 4 classes. De modo geral, a rede neural sem realimentação obteve acurácia aproximada de 97% e precisão aproximada de 87%.

Souza et al. (2015) propuseram a utilização de redes neurais artificiais para prever incêndios florestais no Parque Nacional da Chapada das Mesas, localizado ao sul do estado do Maranhão. A classificação do risco é binarizada. Foi utilizado um *dataset* com dados meteorológicos referentes à 11 meses do ano de 2010. A acurácia da MLP para as amostras de validação foi de 84,79% e a precisão foi de 40,01%.

Goldarag et al. (2016) fizeram um estudo comparando o desempenho de uma rede neural artificial com um modelo de regressão logística para a classificação de risco de incêndios florestais. Enquanto o modelo de regressão logística obteve acurácia de 65,76%, o modelo de rede de múltiplas camadas sem retroalimentação atingiu 93,49% de acurácia.

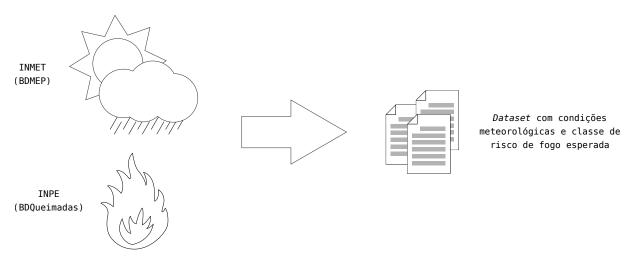
3 MÉTODO DE PESQUISA

Neste capítulo será demonstrado o método de pesquisa para o desenvolvimento do trabalho, compreendendo todos os passos desde a obtenção das informações nos bancos de dados públicos, até a metodologia de análise dos resultados obtidos.

3.1 Obtenção dos dados

Para o desenvolvimento deste trabalho, especificamente para as etapas de treinamento e teste das RNAs, foi utilizada a base de dados do Instituto Nacional de Pesquisas Espaciais (INPE) com os registros de queimadas, e a base de dados do Instituto Nacional de Meteorologia (INMET) com os registros meteorológicos. O cruzamento dos registros de ambos os bancos de dados para uma localidade e intervalo de tempo específicos permite que seja possível construir as amostras para o treinamento e validação dos experimentos, como pode ser visto na Figura 4.

Figura 4 – Diagrama da utilização das duas bases de dados para gerar o dataset



Fonte: Produção do autor

Para obter os dados meteorológicos de uma localidade em um intervalo de tempo específico, é utilizado o "Banco de Dados Meteorológicos para Ensino e Pesquisa" disponível por meio do site do INMET¹. Dentro do sistema, é utilizada a opção de pesquisa da série histórica por "dados diários". A base de dados abrange as seguintes variáveis para o intervalo de tempo selecionado: "temperatura máxima", "temperatura mínima", "insolação", "precipitação", "evaporação do piche", "temperatura média", "média da umidade relativa do ar" e "média da velocidade do vento".

Portal do BDMEP. Disponível em: http://www.inmet.gov.br/portal/index.php?r=bdmep/bdmep. Acesso em: 15 ago. 2019

Após selecionar o município de interesse, é dada a opção para exportar os dados, e assim é gerado um arquivo com valores separados por ponto-e-vírgulas, necessitando que seja posteriormente tratado para que possa ser utilizado para o trabalho.

Para cada dia são feitas duas medições. A primeira é realizada às 00:00 com algumas das variáveis, e a segunda é realizada às 12:00 com o restante delas, tal como demonstrado na Tabela 4. Ambas as medições são referentes ao período de 24 horas anteriores ao horário da medição. Os horários das medições estão registrados seguindo o Tempo Coordenado Universal (UTC). Portanto, para a maioria das cidades do Brasil, o horário das amostras na realidade indicam um instante de 3 horas anteriores ao que está registrado (UTC-03). Ex.: onde na amostra está registrado como 00:00, na realidade está indicando as condições meteorológicas às 21:00 do dia anterior no local.

Porém, deve-se atentar para o fato de que em certos períodos, por conta do horário de verão, as medições ocorrem em horários locais diferentes do convencional. Independente de ocorrer horário de verão em alguma localidade, as medições continuam obedecendo estritamente ao horário fixado ao UTC+00.

Tabela 4 – Definição das variáveis registradas em cada medição

12:00 UTC (09:00 local)	00:00 UTC (21:00 local)
Precipitação	Temperatura Máxima
Temperatura Mínima	Insolação
	Evaporação do Piche
	Temperatura Compensada Média
	Umidade Relativa Média
	Velocidade do Vento Médio

Fonte: Instituto Nacional de Meteorologia

Por meio do *Portal Queimadas* do site do INPE², é disponibilizado o *Banco de Dados de Queimadas* (*BDQueimadas*), em que todas as queimadas identificadas por meio de satélite dentro de um intervalo de tempo em uma região específica, podem ser exportadas para um arquivo de valores separados por vírgulas (CSV).

Ao gerar o arquivo, cada linha é um registro de queimada identificada, e suas colunas demonstram as variáveis daquele local no momento em que ela foi registrada. São elas: "data/hora", "identificação do satélite", "país", "estado", "município", "bioma", "dias de secura", "precipitação", "índice de risco de fogo", "latitude", "longitude", "área industrial" e "potência radiativa do fogo (FRP)".

As variáveis que serão utilizadas deste banco de dados para este trabalho são o "risco de fogo" e os "dias de secura".

Portal Queimadas. Disponível em: http://queimadas.dgi.inpe.br/queimadas/portal. Acesso em: 23 ago. 2019

3.2 Preparação dos dados

Para que não ocorram ruídos na fase de treinamento, é necessária uma análise prévia, correção e/ou remoção de dados inválidos dentre os registros obtidos na etapa anterior.

Considerando que todos os dados oriundos do BDMEP são registrados com fuso-horário UTC+00, então será feita uma correção deles. Isto é necessário principalmente por conta das medições realizadas às 00:00 UTC. A não correção dos horários fará com que os dados de um dia anterior sejam indicados para um dia diferente da realidade. Assim, cada dia deverá ter 2 amostras realizadas, a primeira às 9h locais e a segunda às 21h (caso não esteja ocorrendo o horário de verão).

As estações meteorológicas podem ficar inativas por conta de problemas técnicos, e assim não é realizado o registro das condições naquele momento. Isso provoca a falta de dados que vão desde variáveis específicas até dias inteiros seguidos sem nenhum registro.

Por conta disso, após a correção dos horários, as datas que não possuam as 2 medições realizadas são sumariamente excluídas do *dataset*. Como demonstrado na tabela 4, para cada medição, são obtidas variáveis específicas. A falta de algumas das variáveis dessas medições impossibilita a utilização deste dado para o treinamento da RNA.

Também podem existir casos em que as duas medições foram realizadas para um dia, mas faltam algumas variáveis específicas. Para casos como este, a data apenas é excluída do *dataset* se "temperatura máxima", "umidade relativa média" ou "precipitação" não estiverem presentes.

Feita a exclusão de amostras de datas que não poderão ser utilizadas, é feito um agrupamento de todas as variáveis das duas medições realizadas em um dia, para corresponder também a um mesmo dia no conteúdo do *dataset*.

A preparação das amostras de queimadas consistem primeiramente em verificar se já existem as variáveis "risco de fogo observado" e "dias de secura" no registro de queimada em questão. Caso exista, as variáveis serão mescladas com o *dataset* de dados meteorológicos para a mesma data que o incêndio florestal ocorreu. Caso não exista, será realizado o cálculo manual do "risco de fogo observado" e dos "dias de secura" utilizando a metodologia do INPE.

Como os dados meteorológicos na maioria das vezes começaram a ser registrados muitas décadas atrás, e o programa Queimadas do INPE apenas começou a classificar o índice de Risco de Fogo nos tempos recentes, existe uma grande chance dessas variáveis não estarem presentes para a maioria dos dias filtrados válidos no *dataset* meteorológico, necessitando que estas sejam calculadas no momento do pré-processamento, utilizando as próprias variáveis meteorológicas já disponíveis pelo BDMEP.

Para a construção das amostras caso ocorra a falta dos dados em alguma amostra, será calculado o "Risco de Fogo Observado" para **todo o** *dataset*, inclusive para as outras amostras

em que os dados já existem, sobreescrevendo-os para que não existam inconsistências no reconhecimento do padrão no momento do treinamento.

O cálculo também será feito sem a correção do Risco de Fogo para a detecção de focos de queima. Essa decisão se deve ao fato da correção criar possíveis ruídos no momento do treinamento, ao alterar o índice priorizando a observação factual dos focos de incêndio de dias anteriores, em detrimento da utilização das variáveis meteorológicas da localidade para aquele dia específico.

Com o índice calculado, é possível classificar o risco de incêndio de MÍNIMO a CRÍTICO para cada dia específico (SETZER et al., 2016), abrangendo 5 classes diferentes que poderão ser utilizadas para o treinamento e validação do algoritmo de aprendizado de máquina. A Figura 5 exemplifica os dados de entrada e os dados de saída após a preparação para um dia específico.

Figura 5 – Exemplo da preparação e resultado da junção dos dados

07/09/2006 (queimada registrada) Valores 07/09/2006 (condições meteorológicas) 2006-09-07 16:53:00 Data/Hora 12:00 00:00 (próx. dia) Satélite AQUA_M-T Estação 82564 82564 País Brasil 2006-09-07 12:00:00 Data/Hora 2006-09-08 00:00:00 Estado MARANHA0 Precipitação NULL Município **IMPERATRIZ** Temp. Máxima NULL 36.9 Bioma Cerrado Temp. Mínima 23.5 NULL NULL Dias sem chuva Insolação NULL NULL NULL Precipitação Evapor. Piche 7.7 NULL NULL Risco de Fogo Temp. Média NULL 28.5 Latitude -5.567 Umidade NULL 48.75 -47.472 Longitude Velocid. Vento NULL 1.9 Área Industrial NULL FRP NULL Cálculo dos "dias de secura' Cálculo do RF Classificação do RF Agrupamento das variáveis 07/09/2006 (dados agrupados) Amostra 2391 Valores Data 2006-09-07 Valores Precipitação 0 Precipitação 0.0 Temp. Máxima Temp. Máxima 0.8079470198675497 36.9 Filtragem Umidade 48.75 Umidade 0.1598360655737705 Normalização Dias de Secura 90.72839291156444 0.8676646386297471 Dias de Secura Risco Observado 0.49709939158118543 Classe de Risco Classe de Risco 2

Fonte: Produção do autor

3.2.1 Seleção/extração de características

Com todas as variáveis meteorológicas agrupadas para cada dia, e com as variáveis "risco de fogo observado" e "dias de secura" presentes para cada um dos mesmos dias, todas elas são agrupadas mais uma vez para representarem dados meteorológicos e de queimadas para um dia específico no *dataset*. Como algumas das colunas do *dataset* não são úteis para o aprendizado de máquina, é feita mais uma filtragem das variáveis. Dados como "data", "velocidade do vento" e "temperatura mínima" não são considerados para o cálculo do "Risco de Fogo" do INPE, portanto não serão utilizados. Ao final, apenas estarão presentes as colunas "precipitação", "temperatura máxima", "umidade", "dias de secura" e "classe de risco". Como o objetivo deste trabalho é classificar o risco de incêndio florestal em comparação ao Índice de Risco de Fogo do INPE, as variáveis de "data", "houve incêndio" e "risco observado" em valor real, não serão utilizadas para a etapa de treinamento.

3.2.2 Normalização do dataset

A próxima etapa do pré-processamento consiste na normalização dos valores de cada variável. Será obtido o maior e o menor valor para cada uma, e será utilizada a equação:

$$V_{norm} = \frac{V - V_{min}}{V_{max} - V_{min}} \tag{3.1}$$

Em que V_{norm} é o valor normalizado, V é o valor de entrada não-normalizado, V_{min} é o menor valor para a variável em todo o *dataset* e V_{max} é o maior valor para a variável em todo o *dataset*. Ao final deste cálculo, todas as variáveis de cada uma das amostras terão um valor entre 0 e 1.

3.2.3 Separação dos datasets de treinamento e validação

Para que seja feita a análise estatística do desempenho do modelo da rede neural utilizando o método *hold-out*, é necessária a separação de todo o conjunto do *dataset* entre amostras de treinamento e amostras de validação.

Para este trabalho, 75% das amostras serão utilizadas para o treinamento, e 25% serão utilizados para a validação.

3.2.4 Balanceamento das amostras de treinamento

Ao realizar o treinamento da RNA, o algoritmo ajustará os pesos de cada entrada sucessivamente, para que cada vez mais amostras tenham como classes obtidas as mesmas classes esperadas para elas. Porém, o *dataset* de treinamento pode possuir uma certa quantidade de amostras com a classe esperada, enquanto a quantidade de amostras de outra classe pode ser apenas uma fração deste número. Provocando uma tendência da RNA de enviesar o seu modelo na direção das amostras das classes com maior quantidade (ZHANG et al., 2010).

Para evitar este problema, será realizado o *undersampling* no *dataset* de treinamento. Ou seja, todas as classes de risco de fogo terão consideradas a mesma quantidade de amostras da classe de menor quantidade. Mas antes disto, será feita a aleatorização das posições das amostras, para que não sejam obtidas apenas as primeiras amostras registradas de cada classe (ZHANG et al., 2010). A sequência de passos a serem realizados nesta etapa é demonstrada no Algoritmo 1.

Algoritmo 1: Balanceamento do *dataset* de treinamento Entrada: DatasetInicial = Dataset com todas as amostras [1...n]Amostras $_{ij}$ = Matriz de amostras em que o elemento j pertence à classe iNumClasses = Contagem de classes possíveis Qtd = Contagem de amostras da classe de menor quantidade **Saída:** DatasetFinal = Dataset que será retornado $[1...n_2]$ 1 início aleatorizarPosições(DatasetInicial); 2 $Amostras \leftarrow separarPorClasses(DatasetInicial);$ 3 para $i \leftarrow 1$ até NumClasses faça 4 para $j \leftarrow 1$ até Qtd faça AmostraAtual \leftarrow Amostras_{ij}; 6 DatasetFinal.adicionar(AmostraAtual); fim 8 fim 10 fim

Para o *dataset* de validação, não será realizado o balanceamento. Como a validação busca determinar estatisticamente o desempenho da rede em uma situação hipotética próxima à realidade, sem quantidades de classes previamente definidas, então o balanceamento dos dados para este outro *dataset* não se faz necessário.

3.3 Treinamento

Nesta etapa, primeiramente são definidas as configurações do treinamento da RNA. O modelo que será utilizado consiste em uma Perceptron de Múltiplas Camadas, tendo na camada de entrada as variáveis normalizadas "precipitação", "temperatura máxima", "umidade" e "dias de secura". E terá como saída a classe de 0 até 4 (risco MÍNIMO a CRÍTICO), que representa o retorno do processamento dos valores indicados na camada de entrada.

Além das camadas de entrada e saída, a MLP terá um número de camadas ocultas. Com cada uma delas possuindo uma mesma quantidade de entradas, para que seus pesos sejam sucessivamente ajustados sempre na direção de encontrar uma solução ótima para as saídas

obtidas, em comparação com a saída esperada. A taxa de aprendizado da rede será determinada logo após a abertura do *dataset*. A quantidade de épocas de treinamento será limitada a um número específico, para evitar que o treinamento continue por tempo indeterminado.

Porém, a cada 50 épocas será realizada uma verificação da função de perda(*LOSS*) do modelo da RNA para o *dataset* de validação. Caso a função de perda volte a subir em alguma das verificações, será realizada a *parada antecipada*. Dessa maneira, o treinamento será interrompido, e será salvo o estado da rede na época da última verificação antes daquela que provocou a parada(50 épocas atrás). O algoritmo da etapa de treinamento é demonstrado no Algoritmo 2.

Algoritmo 2: Algoritmo do treinamento com parada antecipada

Entrada: DatasetTreinamento ← Dataset com todas as amostras de treinamento; DatasetValidação ← Dataset com todas as amostras de validação;

MaxEpocas ← Quantidade máxima de épocas;

```
1 início
      ModeloRede \leftarrow (Novo);
2
       ModeloRedeAux \leftarrow ModeloRede;
3
      Overfitting \leftarrow FALSO;
4
      Época ← 0;
5
      LossAnterior \leftarrow 9999;
6
      enquanto Overfitting = FALSO e Época < MaxEpocas faça
7
          ajustar(ModeloRede, DatasetTreinamento);
8
          LossAtual ← calcularFuncaoPerda(DatasetValidação);
          se mod(Época, 50) = 0 então
10
              se LossAtual > LossAnterior então
11
                  Overfitting \leftarrow VERDADEIRO;
12
                  ModeloRede \leftarrow ModeloRedeAux;
13
              senão
14
                  ModeloRedeAux \leftarrow ModeloRede;
15
              fim
16
          fim
17
          LossAnterior \leftarrow LossAtual;
18
          Época ← Época + 1;
19
      fim
20
      salvar(ModeloRede);
21
22 fim
```

3.4 Análise do desempenho da RNA

Ao término da etapa de treinamento do modelo da RNA, é realizada uma verificação dos dados gerais com as amostras de treinamento e validação. Informações como função de perda, acurácia e precisão tanto do modelo geral quanto de cada uma das classes serão analisadas com base em dados gravados a cada época do treinamento.

3.5 Análise de acertos em relação à incêndio registrado

Com o modelo da RNA pronto, as amostras de um intervalo específico de tempo serão processadas utilizando este modelo de aprendizado de máquina, e também serão processadas utilizando o modelo matemático descrito pelo INPE. Após este processamento, será analisada a taxa de acerto de ambos para a classificação em relação à presença ou ausência de incêndio florestal.

Para as classes de risco 0 e 1, a taxa de acerto será definida pela **ausência** de fogo nas amostras em que foram indicadas estas mesmas classes de risco. Enquanto para as classes de risco 2, 3 e 4, a taxa de acerto será definida pela **presença** de fogo nas amostras do intervalo de tempo determinado.

Esta análise será realizada para ambos os modelos (INPE e RNA) e será feita a comparação da taxa de acerto entre eles. Ela é baseada na página de "Análise de Acertos" disponível no Portal Queimadas do INPE³. Neste portal é apresentado somente a porcentagem de como havia sido classificado o risco quando existiram registros de queimadas. Neste trabalho, será apresentado também o que foi classificado quando não foi detectada alguma queimada.

Além disso, a página do INPE agrupa todos os registros de queimadas dos dias classificados por regiões do Brasil, enquanto neste trabalho serão utilizados os registros de um único município.

³ http://prodwww-queimadas.dgi.inpe.br/queimadas/analise-risco-fogo/

4 EXPERIMENTO

Neste capítulo será demonstrado o experimento a ser realizado, com todas as suas variáveis desde o pré-processamento até a etapa de treinamento. O experimento foi realizado em um computador de mesa com sistema operacional Debian 9 de 64 bits, processador AMD FX-6300 de 3,5GHz e memória RAM de 8 GB. Todos os códigos para a execução foram escritos na linguagem Java, utilizando o ambiente de desenvolvimento integrado Eclipse, versão 2019-03. E para o treinamento da rede neural, foi utilizada a biblioteca de código aberto *DeepLearning4J*(DL4J)¹.

A localidade escolhida para este experimento foi o município de Imperatriz, localizado no sudoeste do estado do Maranhão. A escolha desta área em particular se deve principalmente ao fato dela estar inserida dentro da chamada "Amazônia Legal", área que foi delimitada pelo Estado brasileiro para que algumas políticas públicas — inclusive para a prevenção de queimadas — sejam priorizadas (IBGE, 2006).

Também foi escolhido por estar próximo à linha do Equador, tal como pode ser visto na Figura 6, assim possuindo grande variação no regime de precipitação, com chuvas predominantes entre os meses de outubro e maio, ao mesmo tempo em que durante todo o ano as temperaturas máximas diárias são altas. Além disso, a cidade se localiza em uma região de transição entre Amazônia e Cerrado.

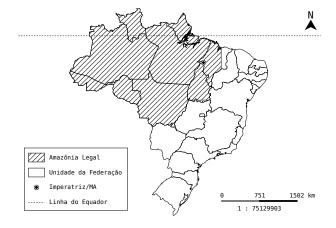


Figura 6 – Localização de Imperatriz/MA na "Amazônia Legal"

Fonte: Produção do autor a partir de dados do IBGE (2012)

DeepLearning4J. Disponível em: https://deeplearning4j.org/. Acesso em: 20 set. 2019

4.1 Dados de entrada

Para este município em particular, foram obtidos do BDMEP dados meteorológicos a partir do dia 01/06/1961 até o dia 31/12/2018, contendo no total 35.861 registros meteorológicos, entre medições realizadas às 00:00 UTC e 12:00 UTC. E a partir do BDQueimadas, foram obtidos 6150 registros de queimadas, entre os dias 01/01/1999 e 06/12/2018.

Nos dados do BDMEP, após a correção dos horários, filtragem e agrupamento por datas, a tabela de dados meteorológicos ficou com 15467 registros validados.

No caso do BDQueimadas, foi necessário o cálculo manual do "risco de fogo" e "dias de secura" de cada registro, pois o cálculo para este município apenas começou a ser realizado em abril/2017, e consequentemente existiriam apenas 373 registros. Durante o cálculo do risco de fogo, foi utilizada como constante A o valor referente à classe "Savana; Caatinga Aberta", que é igual a 3,0. A localidade escolhida possui esse tipo de vegetação segundo o IBGE (2004). Como o cálculo manual foi feito, foram validados e classificados 6.030 registros.

Para os dias existentes no BDMEP, mas inexistentes do BDQueimadas, também foi realizado o cálculo manual. E após o agrupamento de todos os dados, foi construído um *dataset* com 15.467 amostras.

Foi realizada a normalização do *dataset*, e ele foi dividido em 11.601 amostras de treinamento e 3.866 amostras de validação. Após a aleatorização e balanceamento do dataset de treinamento, foi obtido um *dataset* final com 235 amostras, sendo 47 amostras para cada classe.

4.2 Treinamento

Com os *datasets* de treinamento e validação preparados para o treinamento, é possível fazer a configuração da rede neural. Neste experimento, a MLP terá 4 camadas ocultas, com cada uma possuindo 5 entradas. A taxa de aprendizado será configurada para 0,1. E a quantidade de épocas será limitada a 100 mil. A cada época serão registrados os dados da função de perda para ambos os *datasets*, e dados de acurácia e precisão para cada classe, assim como a média.

5 RESULTADOS

Neste capítulo serão apresentados os resultados do experimento realizado, e será demonstrada a taxa de acerto da RNA ao classificar o risco de fogo, utilizando a metodologia do INPE para medir a taxa de acerto para os dados de risco de fogo obtidos no BDQueimadas.

5.1 Treinamento

A etapa de treinamento do modelo da RNA foi finalizada com 100 mil épocas. Portanto, até esta época, a função de perda continuava decrementando, como pode ser vista na Figura 7. Indicando que a rede ainda não havia atingido sua menor função de perda.

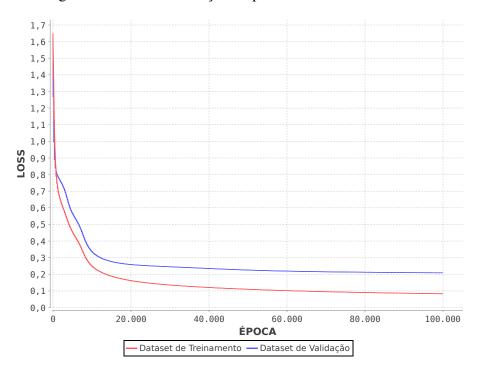


Figura 7 – Gráfico da função de perda durante o treinamento

Fonte: Produção do autor

Ao final do treinamento, o modelo da RNA alcançou 93,99% de acurácia e 87,19% de precisão para as 3866 amostras de validação, como pode ser visto nas Figuras 8 e 9. Para as classes 0, 1, 2 e 3, a rede obteve 99,28%, 90,19%, 91,08% e 94,56% de precisão, respectivamente. Mas para a classe 4, a rede teve apenas 60,87% de precisão, como pode ser visto na Figura 10.

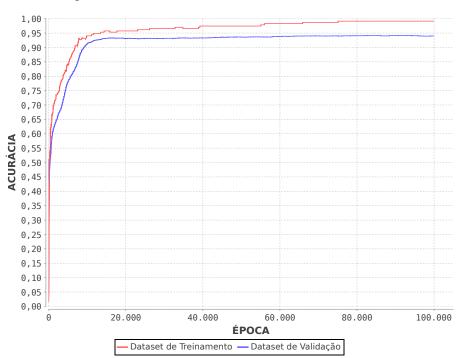


Figura 8 – Gráfico da acurácia durante o treinamento

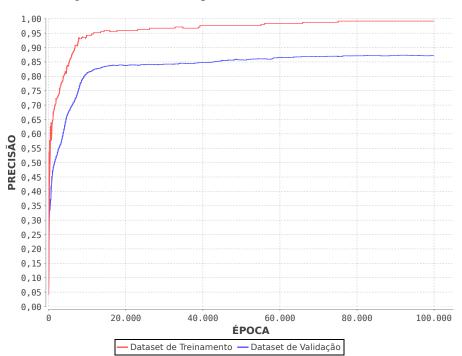


Figura 9 – Gráfico da precisão durante o treinamento

Fonte: Produção do autor

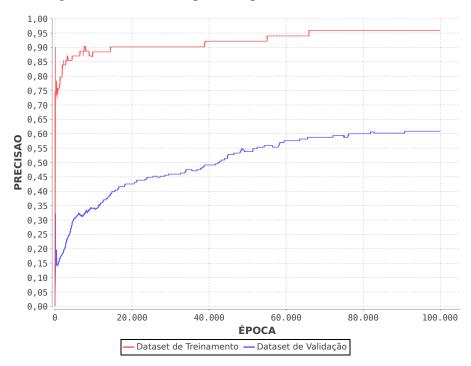


Figura 10 – Gráfico da precisão para a classe 4 (MÁXIMO)

5.2 Análise de acerto para incêndios registrados

Para realizar esta modalidade de análise de acertos, foram utilizadas todas as amostras válidas referentes ao ano de 2018, totalizando 215 registros. De modo geral, para os dias de incêndios registrados, a MLP obteve uma taxa de acerto menor em comparação ao modelo matemático do INPE.

Para os dias em que o risco de incêndio foi classificado como 0 ou 1 utilizando o modelo matemático, em 83,61% destes dias a queimada foi inexistente de fato, como pode ser visto na Figura 11a. Para os dias em que o risco foi classificado com estas mesmas classes utilizando o modelo da RNA, em 80,33% deles não ocorreram incêndio, como pode ser visto na Figura 12a. Para os dias classificados como risco 2, 3 ou 4 pelo modelo matemático, em 51,61% deles realmente ocorreram incêndio, enquanto a MLP classificou da mesma maneira em 47,31% dos dias em que queimadas foram presentes, como demonstrado pelas Figuras 11b e 12b.

Figura 11 – Gráficos com a indicação da presença de fogo referente ao agrupamento de classes calculadas pelo INPE

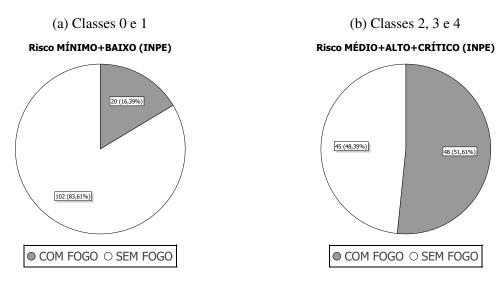
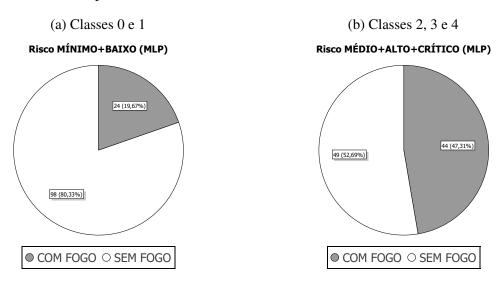


Figura 12 – Gráficos com a indicação da presença de fogo referente ao agrupamento de classes calculadas pela MLP



Fonte: Produção do autor

Nota-se que ao realizar a análise do desempenho dos dois modelos com cada classe separada, em algumas delas os resultados são idênticos. As Figuras 13 e 14 demonstram graficamente a semelhança na quantidade classificada pelos dois modelos. Ambos os modelos classificaram 66 dias como risco 0, e em 63 deles realmente não existiram incêndio, como visto nas Tabelas 5 e 6, obtendo uma taxa de acerto de 95,45%.

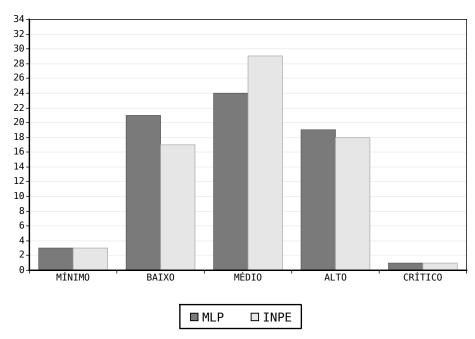


Figura 13 – Quantidade de amostras para cada classe quando houve fogo

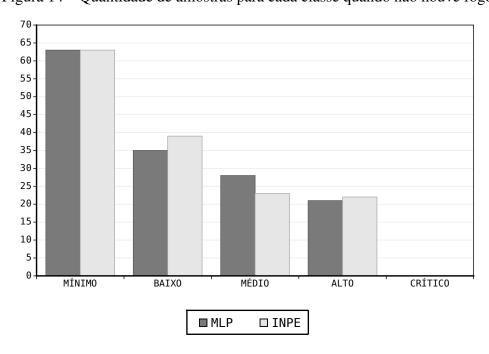


Figura 14 – Quantidade de amostras para cada classe quando não houve fogo

Fonte: Produção do autor

Para a classe 1, 39 das amostras corresponderam à dias em que realmente não existiu incêndio para o modelo do INPE, enquanto a MLP classificou como este risco, 35 de 56 amostras.

Dessa maneira, a taxa de acerto do modelo do INPE para a classe 1 foi de aproximadamente 69,64%, enquanto o da MLP foi de 62,5%.

Para a classe 2, 29 de 52 amostras foram referentes à dias com queimada pelo modelo de classificação do INPE, enquanto para a MLP esse número foi de 24 de 52 amostras. A taxa de acerto do modelo do INPE para esta classe é de 55,77%, enquanto a taxa de acerto do modelo MLP foi de 46,15%.

Para a classe 3, 18 de 40 amostras corresponderam à dias com queimada utilizando o modelo do INPE, enquanto o modelo MLP essa quantidade foi de 19 de 40 amostras. Assim demonstrando taxa de acerto de 45% para o modelo do INPE, e 47,5% de taxa de acerto para o modelo MLP. Neste caso, o modelo MLP obteve taxa de acerto maior que o modelo do INPE.

Tabela 5 – Verificação INPE

Classe	Com Fogo	Sem Fogo	Total	Acerto(%)
0	3	63	66	95,45
1	17	39	56	69,64
2	29	23	52	55,77
3	18	22	40	45
4	1	0	1	100
0+1	20	102	122	83,61
2+3+4	48	45	93	51,61

Fonte: Produção do autor

Tabela 6 – Verificação MLP

Classe	Com Fogo	Sem Fogo	Total	Acerto(%)
0	3	63	66	95,45
1	21	35	56	62,5
2	24	28	52	46,15
3	19	21	40	47,5
4	1	0	1	100
0+1	24	98	122	80,33
2+3+4	44	49	93	47,31

Fonte: Produção do autor

E para a classe 4, pelo fato de apenas 1 amostra ser prevista por ambos os modelos (INPE e MLP), nesta única amostra realmente ocorreu incêndio florestal. Sendo assim, a classe 4 obteve 100% de taxa de acerto para ambos os modelos.

Ao calcular a média das taxas de acerto de todas as classes, o modelo do INPE possui aproximadamente 73,17% de acerto, enquanto o modelo da MLP possui 70,32% de taxa de acerto.

6 CONCLUSÃO

Este trabalho propôs a utilização de conceitos de redes neurais artificiais para a classificação de risco de incêndios florestais, utilizando grandes quantidades de dados disponíveis por meio do Banco de Dados Meteorológicos para Ensino e Pesquisa (BDMEP) e Banco de Dados de Queimadas (BDQueimadas).

A partir do desenvolvimento deste trabalho, conclui-se que o treinamento de uma RNA utilizando as variáveis meteorológicas e de queimadas possibilita a classificação de risco de incêndios florestais para o município de Imperatriz/MA. O fato da RNA realizar seu treinamento com base em saídas esperadas que foram classificadas anteriormente pelo modelo do INPE, não permite que sua taxa de acerto seja superior à este mesmo modelo matemático do INPE. Em outras palavras, a rede neural tende a reproduzir a classificação da mesma maneira que o INPE, portanto não é possível ter taxa de acerto maior do que o modelo matemático descrito por Setzer et al. (2016).

Porém, por meio da análise do desempenho da RNA, é possível observar que ainda existe a possibilidade de aumento na sua taxa de acerto. Como o treinamento foi finalizado em seu limite de 100 mil épocas, pode-se assumir que a rede ainda não havia atingido seu estado com menor função de perda.

Além disso, a classe 4 teve um aumento mais lento na sua precisão para as amostras de validação durante o treinamento. O que pode causar uma baixa taxa de acerto para as amostras a serem classificadas como risco MÁXIMO.

Como trabalho futuro, sugere-se a adição ou exclusão de variáveis para aumentar especificamente a precisão do modelo da rede neural para a classe 4. Também é sugerida a utilização de outros tipos de RNA para a classificação. Além disso, sugere-se o treinamento e análise da RNA para a classificação de incêndio sem utilizar os dados do índice RF do INPE. Dessa maneira será possível obter taxa de acerto maior que o modelo matemático do INPE para incêndios registrados.

REFERÊNCIAS

CNBC. Deadly California Wildfire Now 100% Contained After Scorching 154,000 Acres. [S.l.]: CNBC, 2018. https://www.cnbc.com/2018/11/25/ deadly-california-wildfire-now-100percent-contained.html>. Acesso em: 05 mar. 2019.

DEVELOPERS, G. Descending into ml: Training and loss. In: _____. *Machine Learning Crash Course*. [S.l.]: Google Developers, 2019. Acesso em: 05 set. 2019.

FONSECA-MORELLO, T. et al. Fires in Brazilian Amazon: Why Does Policy Have A Limited Impact? *Ambiente & Sociedade*, SCIELO, v. 20, p. 19 – 38, 12 2017. ISSN 1414-753X. Disponível em: ">http://www.scielo.br/scielo.php?script=sci_arttext&pid=S1414-753X2017000400019&nrm=iso>">http://www.scielo.br/scielo.php?script=sci_arttext&pid=S1414-753X2017000400019&nrm=iso>">http://www.scielo.br/scielo.php?script=sci_arttext&pid=S1414-753X2017000400019&nrm=iso>">http://www.scielo.br/scielo.php?script=sci_arttext&pid=S1414-753X2017000400019&nrm=iso>">http://www.scielo.br/scielo.php?script=sci_arttext&pid=S1414-753X2017000400019&nrm=iso>">http://www.scielo.br/scielo.php?script=sci_arttext&pid=S1414-753X2017000400019&nrm=iso>">http://www.scielo.br/scielo.php?script=sci_arttext&pid=S1414-753X2017000400019&nrm=iso>">http://www.scielo.br/scielo.php?script=sci_arttext&pid=S1414-753X2017000400019&nrm=iso>">http://www.scielo.br/scielo.php?script=sci_arttext&pid=S1414-753X2017000400019&nrm=iso>">http://www.scielo.br/scielo.php?script=sci_arttext&pid=S1414-753X2017000400019&nrm=iso>">http://www.scielo.br/scielo.php?script=sci_arttext&pid=S1414-753X2017000400019&nrm=iso>">http://www.scielo.php?script=sci_arttext&pid=S1414-753X2017000400019&nrm=iso>">http://www.scielo.php?script=sci_arttext&pid=S1414-753X2017000400019&nrm=iso>">http://www.scielo.php?script=sci_arttext&pid=S1414-753X2017000400019&nrm=iso>">http://www.scielo.php?script=sci_arttext&pid=S1414-753X2017000400019&nrm=iso>">http://www.scielo.php?script=sci_arttext&pid=S1414-753X2017000400019&nrm=iso>">http://www.scielo.php?script=sci_arttext&pid=S1414-753X2017000400019&nrm=iso>">http://www.scielo.php?script=sci_arttext&pid=S1414-753X2017000400019&nrm=iso>">http://www.scielo.php?script=sci_arttext&pid=S1414-753X2017000400019&nrm=iso>">http://www.scielo.php?script=sci_arttext&pid=S1414-753X2017000400019&nrm=iso>">http://www.scielo.php?sci_arttext&pid=S1414-753X2017000400019&nrm=iso>">ht

GOLDARAG, Y. J.; MOHAMMADZADEH, A.; ARDAKANI, A. S. Fire risk assessment using neural network and logistic regression. *Journal of the Indian Society of Remote Sensing*, Springer, v. 44, n. 6, p. 885–894, 2016.

HAYKIN, S. S. *Neural Networks And Learning Machines*. Third. Upper Saddle River, NJ: Pearson Education, 2009.

IBGE. *Mapa de Vegetação do Brasil*. [S.l.]: Instituto Brasileiro de Geografia e Estatística, 2004. ftp://geoftp.ibge.gov.br/informacoes_ambientais/vegetacao/mapas/brasil/vegetacao.pdf. Acesso em: 20 set. 2019.

IBGE. *Mapa Temático da Amazônia Legal*. [S.l.]: Instituto Brasileiro de Geografia e Estatística, 2006. https://mapas.ibge.gov.br/tematicos/amazonia-legal.html. Acesso em: 25 set. 2019.

KOHAVI, R. et al. A study of cross-validation and bootstrap for accuracy estimation and model selection. In: MONTREAL, CANADA. *Ijcai*. [S.l.], 1995. v. 14, n. 2, p. 1137–1145.

LALL, S.; MATHIBELA, B. The application of artificial neural networks for wildfire risk prediction. In: 2016 International Conference on Robotics and Automation for Humanitarian Applications (RAHA). [S.l.: s.n.], 2016. p. 1–6.

MCBRIDE, A.; GUTIERREZ, M.; ASIMOV, N. 29 Dead In Camp Fire As Firefighters Make Gains On Butte County Blaze. [S.l.]: San Francisco Chronicle, 2018. https://www.sfgate.com/california-wildfires/article/Butte-County-fire-only-grows-slightly-as-13381900.php. Acesso em: 05 mar. 2019.

PARMAR, R. *Common Loss functions in machine learning*. 2018. https://towardsdatascience.com/common-loss-functions-in-machine-learning-46af0ffc4d23. Acesso em: 05 set. 2019.

PROVOST, F.; KOHAVI, R. Glossary of terms. *Journal of Machine Learning*, v. 30, n. 2-3, p. 271–274, 1998.

RUSSELL, S. J.; NORVIG, P. Artificial Intelligence. 3. ed. [S.l.]: Prentice Hall, 2010.

SAYAD, Y. O.; MOUSANNIF, H.; MOATASSIME, H. A. Predictive modeling of wildfires: A new dataset and machine learning approach. *Fire Safety Journal*, v. 104, p. 130–146, 2019.

- SETZER, A. W.; SISMANOGLU, R. A.; MARTINS, G. *Metodologia do Cálculo do Risco de Fogo do Programa Queimadas do INPE-Versão 10*. 2016. http://urlib.net/8JMKD3MGP3W34P/3PNK9TH.
- SILVA, I. D. B.; BARROS, L. C.; MEYER, J. F. C. A. Modelagem de risco e perigo de incêndios florestais com simulações de propagação. *Proceeding Series of the Brazilian Society of Computational and Applied Mathematics*, v. 6, n. 2, 2018.
- SISMANOGLU, R. A.; SETZER, A. W. Avaliação da previsão do risco de fogo desenvolvida no *INPE aplicando o modelo ETA*: Análise objetiva e comparativa entre duas versões para 2003. 2004.
- SOUZA, F. T. de; KOERNER, T.; CHLAD, R. A data-based model for predicting wildfires in chapada das mesas national park in the state of maranhão. *Environmental Earth Sciences*, v. 74, 05 2015.
- WIKIMEDIA. Loss function. In: ____. *Wikipedia, The Free Encyclopedia*. [S.l.]: Wikimedia, 2019. Acesso em: 05 set. 2019.
- WOLPERT, D. H. Stacked generalization. *Neural Networks*, v. 5, n. 2, p. 241 259, 1992. ISSN 0893-6080. Disponível em: http://www.sciencedirect.com/science/article/pii/S0893608005800231.
- YADAV, S.; SHUKLA, S. Analysis of k-fold cross-validation over hold-out validation on colossal datasets for quality classification. In: 2016 IEEE 6th International Conference on Advanced Computing (IACC). [S.l.: s.n.], 2016. p. 78–83.
- ZHANG, Y. P.; ZHANG, L. N.; WANG, Y. C. Cluster-based majority under-sampling approaches for class imbalance learning. In: *2010 2nd IEEE International Conference on Information and Financial Engineering*. IEEE, 2010. Disponível em: https://doi.org/10.1109/icife.2010.5609385.



ANEXO A - FÓRMULAS DOS FATORES DE PRECIPITAÇÃO

$$fp1 = e^{-0.14 \cdot prec1} \tag{A.1}$$

$$fp2 = e^{-0.07 \cdot (prec2 - prec1)}$$
 (A.2)

$$fp3 = e^{-0.04 \cdot (prec3 - prec2)} \tag{A.3}$$

$$fp4 = e^{-0.03 \cdot (prec4 - prec3)} \tag{A.4}$$

$$fp5 = e^{-0.02 \cdot (prec5 - prec4)} \tag{A.5}$$

$$fp6a10 = e^{-0.01 \cdot (prec10 - prec5)}$$
 (A.6)

$$fp11a15 = e^{-0.008 \cdot (prec15 - prec10)}$$
 (A.7)

$$fp16a30 = e^{-0.004 \cdot (prec30 - prec15)}$$
 (A.8)

$$fp31a60 = e^{-0.002 \cdot (prec60 - prec30)}$$
 (A.9)

$$fp61a90 = e^{-0.001 \cdot (prec90 - prec60)}$$
 (A.10)

$$fp91a120 = e^{-0.0007 \cdot (prec120 - prec90)}$$
 (A.11)

Em que prec1 é a precipitação acumulada para o dia anterior à data de interesse, prec2 é o acumulado do dia anterior até 2 dias anteriores, prec3 do dia anterior até 3 dias anteriores, prec4 do dia anterior até 4 dias anteriores, prec5 do dia anterior até 5 dias anteriores, prec10 do dia anterior até 10 dias anteriores, e assim por diante (SETZER et al., 2016).