

Universidade Federal de Ouro Preto

Núcleo de Geotecnia

Programa de Pós-Graduação em Geotecnia
PPGEO

Tese

**ANÁLISE MULTICRITÉRIO E
APRENDIZADO DE MÁQUINA
APLICADOS NA PREDIÇÃO DO
POTENCIAL ESPELEOLÓGICO
DA REGIÃO DO PARQUE
NACIONAL SERRA DO
GANDARELA, QUADRILÁTERO
FERRÍFERO/MG**

Iraydes Tálita de Sena Nola

Ouro Preto
2022



UFOP

Iraydes Tálita de Sena Nola

**ANÁLISE MULTICRITÉRIO E APRENDIZADO DE
MÁQUINA APLICADOS NA PREDIÇÃO DO
POTENCIAL ESPELEOLÓGICO DA REGIÃO DO
PARQUE NACIONAL SERRA DO GANDARELA,
QUADRILÁTERO FERRÍFERO/MG**

Texto apresentado ao Programa de Pós-Graduação em Geotecnia do Núcleo de Geotecnia da Escola de Minas da Universidade Federal de Ouro Preto, como parte integrante dos requisitos para obtenção do título de **Doutor em Geotecnia.**

Orientador: Prof. Dr. Luís de Almeida Prado Bacellar

**OURO PRETO
2022**

SISBIN - SISTEMA DE BIBLIOTECAS E INFORMAÇÃO

N786a Nola, Iraydes Tálita de Sena.

Análise multicritério e aprendizado de máquina aplicados na predição do potencial espeleológico da região do Parque Nacional Serra do Gandarela, Quadrilátero Ferrífero/MG. [manuscrito] / Iraydes Tálita de Sena Nola. - 2022.
138 f.

Orientador: Prof. Dr. Luis de Almeida Prado Bacellar.

Tese (Doutorado). Universidade Federal de Ouro Preto. Núcleo de Geotecnia da Escola de Minas. Programa de Pós-Graduação em Geotecnia.

Área de Concentração: Geotecnia.

1. Cavidades Ferríferas. 2. Aprendizado do computador. 3. Redes neurais (Computação). 4. Algoritmos computacionais - Florestas Aleatórias. I. Bacellar, Luis de Almeida Prado. II. Universidade Federal de Ouro Preto. III. Título.

CDU 624.13

Bibliotecário(a) Responsável: Maristela Sanches Lima Mesquita - CRB-1716



FOLHA DE APROVAÇÃO

Iraydes Tálita de Sena Nola

Análise multicritério e aprendizado de máquina aplicados na predição do potencial espeleológico da região do Parque Nacional Serra do Gandarela, Quadrilátero Ferrífero/MG

Tese apresentada ao Programa de Pós-Graduação em Geotecnia do Núcleo de Geotecnia da Escola de Minas da Universidade Federal de Ouro Preto, como parte integrante dos requisitos para obtenção do título de **Doutor em Geotecnia**

Aprovada em 03 de fevereiro de 2022

Membros da banca

Prof. Dr. Luis de Almeida Prado Bacellar - Orientador (Universidade Federal de Ouro Preto)
Prof. Dr. César Falcão Barela - (Universidade Federal de Ouro Preto)
Profª. Drª. Maria Giovana Parizzi - (Universidade Federal de Minas Gerais)
Prof. Dr. Allan Erlikhaman - (Centro Federal de Educação Tecnológica de Minas Gerais)
Dr. Marcelo Roberto Barbosa (Vale S.A.)

O Prof. Dr. Luis de Almeida Prado Bacellar, orientador do trabalho, aprovou a versão final e autorizou seu depósito no Repositório Institucional da UFOP em 08/08/2022.



Documento assinado eletronicamente por **Luis de Almeida Prado Bacellar, COORDENADOR(A) DE CURSO DE PÓS-GRADUAÇÃO EM GEOTECNIA**, em 08/08/2022, às 13:58, conforme horário oficial de Brasília, com fundamento no art. 6º, § 1º, do [Decreto nº 8.539, de 8 de outubro de 2015](#).



A autenticidade deste documento pode ser conferida no site http://sei.ufop.br/sei/controlador_externo.php?acao=documento_conferir&id_orgao_acesso_externo=0, informando o código verificador **0375865** e o código CRC **D9BBDD8E**.

À minha querida e amada
Vovó Tina (*in memoriam*),
pelo orgulho que sempre
teve de mim.

AGRADECIMENTOS

Ao Prof. Dr. Luís Bacellar pela orientação dedicada, pela amizade e pelas trocas de conhecimento que foram essenciais no desenvolvimento deste trabalho.

À Universidade Federal de Ouro Preto, pela oportunidade de realização do doutoramento. E ao Conselho Nacional de Desenvolvimento Científico e Tecnológico (CNPq), pela bolsa concedida.

Ao Núcleo de Geotecnia da Escola de Minas, igualmente ao quadro de docentes e demais funcionários, pela infraestrutura e auxílios técnico-administrativos oferecidos, em especial ao Prof. Hernani Lima, ao Prof. Romero Gomes, e ao Prof. Cesar Barella.

Aos amigos do NUGEO: Christ Paria, Maria Isabel Martinez, Mateus Oliveira, Matheus Cardoso, Vera Tavares, Renata Delício e Yesser Rodriguez.

Aos amigos que a Geotecnia me presenteou: Christopher Fonseca, João Manoel, Jussara Michette, Ligia Sampaio e Thaise Morais.

Aos professores da Universidade Federal de Viçosa que me ajudaram a retirar cada pedra que surgia no meio do caminho ao longo da pesquisa, em especial: Prof. Dr. Rodolpho Neves (Dep. Engenharia Elétrica), Profa. Dra. Nilcilene (Dep. Engenharia Civil), Prof. Dr. José Ivo Ribeiro (Dep. Estatística), Prof. Dr. Cosme Damião (Dep. Biologia Geral).

Ao meu querido Roberto, por transformar os momentos escuros em claros como um dia ensolarado de verão. Seu apoio geotécnico, ideias, brincadeiras, carinho e atenção, fizeram toda a diferença.

Às amigas tão valiosas e verdadeiras das queridas Alessandra Souza, Flávia Cristina, Flávia de Faria, Karin Voll, Marina Carvalho.

A minha família Sena e Nola (tios, primos e agregados) pelo amor que vocês têm por mim e porque sei que as minhas realizações são também de cada um de vocês.

À família Santos por me acolherem com tanto carinho e me proporcionarem momentos de descanso e descontração.

E por fim, agradeço a Deus por este trabalho e pela magnífica família que Ele me deu. Obrigada aos meus pais, Irayde Sena e Luiz Nola, aos meus irmãos Luiz Henrique, Luciana e Ana Paula e aos cunhados Cléria e Marcello. Obrigada pelas inúmeras orações, Mamãe, Deus te abençoe.

“... Ninguém ignora tudo.
Ninguém sabe tudo.
Todos nós sabemos alguma coisa.
Todos nós ignoramos alguma coisa.
Por isso aprendemos sempre.”

Paulo Freire

RESUMO

No Brasil, cavidades ferríferas estão geneticamente e geograficamente associadas aos depósitos minerais explotáveis de ferro. Essas cavidades possuem relevância espeleológica, arqueológica, paleontológica e biológica, sendo ambientalmente e legalmente protegidas. Assim, para um melhor planejamento da exploração e licenciamento ambiental torna-se necessário estudar a gênese e o desenvolvimento das cavidades ferríferas buscando preservá-las sem impedir o avanço da mineração. Esse tema é complexo, pouco difundido e os trabalhos que buscam alternativas para prever a ocorrência dessas cavidades são escassos. Essa demanda justifica o desenvolvimento de pesquisas e produtos capazes de auxiliar tomadores de decisão, planejadores e autoridades competentes para suporte na definição de locais-alvo de prospecção espeleológica de campo. Nesta pesquisa foi avaliada a capacidade preditiva de técnicas utilizando lógica *fuzzy* e processo hierárquico analítico, além de técnicas de aprendizado de máquina na avaliação da suscetibilidade a ocorrência de cavidades ferríferas. Para isso, foram selecionadas as variáveis: razão de óxido de ferro, declividade, índice de vegetação por diferença normalizada (NDVI), curvatura horizontal, curvatura vertical, intensidade de lineamentos e HAND (*height above nearest drainage*). Para a validação dos resultados foi utilizado um inventário de cavidades previamente levantado. Essas variáveis foram produzidas a partir do processamento de dados geoespaciais de uma região da Serra do Gandarela, Minas Gerais, Brasil. A primeira técnica aplicada, utilizando lógica *fuzzy* e processo hierárquico analítico (AHP), foi implementada e os resultados mostraram um desempenho satisfatório do mapa produzido em prever áreas favoráveis a ocorrência de cavidades ferríferas, apresentando AUC (*the area under the curve*) em torno de 0,85. Na sequência foi aplicada a técnica de aprendizado de máquina denominada Florestas Aleatórias. Esta técnica foi a que se mostrou mais eficaz no estudo da suscetibilidade a ocorrência de cavidades ferríferas, apresentando AUC igual a 0,957. De forma complementar foi utilizada a técnica de Redes Neurais Artificiais, cujos resultados indicaram que o modelo aplicado não possui capacidade de generalização, apesar de seu desempenho ter sido considerado satisfatório durante o treinamento. Espera-se que os passos metodológicos detalhados neste estudo incentivem os planejadores e tomadores de decisão de empreendimentos de mineração e agências ambientais a adotá-los com o objetivo de mapear a suscetibilidade à ocorrência de cavidades ferríferas totalmente com base na aquisição e análise de dados remotos.

Palavras-chave: Cavidades Ferríferas; Processo Hierárquico Analítico; Aprendizado de Máquina; Florestas Aleatórias; Redes Neurais Artificiais.

ABSTRACT

In Brazil, iron formation caves are genetically and geographically associated with the exploitable mineral deposits of iron. These caves have speleological, archaeological, paleontological and biological relevance, and are environmentally and legally protected. Thus, for a better exploitation planning and environmental licensing it is necessary to study the genesis and development of the iron formation caves to preserve them without obstructing the advance of the mining industry. This is a very complex topic, little researched, and few papers evaluate alternatives for predicting the occurrence of these types of caves. This gap justifies the development of research capable of assisting decision makers, planners and competent authorities to support the definition of target sites for speleological field prospection. In this research, the ability of fuzzy logic, hierarchical analytical process, and machine learning to evaluate the susceptibility to the occurrence of ferruginous caves was evaluated. For this, the variables: iron oxide ratio, slope, normalized difference vegetation index (NDVI), horizontal curvature, vertical curvature, lineament intensity, and height above nearest drainage (HAND) were selected. A previously surveyed cavity inventory was used to validate the results. A database of caves was used to validate the results. These variables were obtained by processing of geospatial data from a region of the Gandarela Range, Minas Gerais, Brazil. The fuzzy logic and analytic hierarchy process were implemented and the results showed a satisfactory performance of the map produced in predicting areas favorable for the occurrence of iron ore caves, with an AUC (the area under the curve) of approximately 0.850. The Random Forest technique was also implemented and proved to be even more effective in studying the susceptibility to the occurrence of ferruginous caves, showing an AUC equal to 0.957. In a complementary manner, the Artificial Neural Networks technique was used, the results of which indicated that the model applied does not have generalization capabilities, despite its performance having been considered satisfactory during training. It is expected that the methodology presented in this study will encourage planners and decision makers of mining enterprises and environmental agencies to adopt them with the aim of mapping the susceptibility to the occurrence of iron formation caves entirely based on remote data.

Keywords: Iron Formation Caves; Analytic Hierarchy Process; Machine Learning; Random Forest; Artificial Neural Network.

LISTA DE FIGURAS

Figura 1: Caverna ferrífera, encontrada na Serra de Carajás. Foto cedida por Georgete Dutra.	22
Figura 2: Modelo de gênese e evolução criado para a Caverna de Junção/MG, que ilustra a formação das cavidades em ambiente vadoso (Calux, 2013).	24
Figura 3: Perfil síntese de uma caverna ferrífera com as principais feições morfológicas. Fonte: Piló et al., (2015).	27
Figura 4: Estrutura hierárquica do processo de AHP.	34
Figura 5: Fases do processo de busca de conhecimento utilizando mineração de dados.	38
Figura 6: Exemplo do processo de Floresta Aleatória.	42
Figura 7: Neurônio artificial simples (Nola, 2015)	47
Figura 8: Representação de uma rede simples, tipo MLP, com uma camada intermediária.	49
Figura 9: Sentido de transmissão do sinal no processo de aprendizagem usando o algoritmo <i>backpropagation</i>	50
Figura 10: Diagrama representativo do processo de elaboração da pesquisa, por etapas.	53
Figura 11: Location map of the study area indicating elevation and the geological units.	63
Figure 12: A) Weathering profile of an itabirite outcrop; B) Iron formation cave in FQ with typical irregular morphology; C) Schematic geological cross-section of a typical iron formation cave.	65
Figure 13: Flow chart for iron formation caves susceptibility assessment.	69
Figure 14: Data layers: (a) iron oxide ratio; (b) slope gradient; (c) plan curvature; (d) NDVI; (e) profile curvature; (f) lineaments intensity; and (g) HAND.	74
Figure 15: Susceptibility map of the study area produced by Analytical Hierarchy Process.	77
Figure 16: Receiver operating characteristic (ROC) curve analysis: (a) success rate curve using the training dataset; (b) predicted curve using the validation dataset.	78
Figure 17: Location of the study area.	83
Figure 18: Lithological map and location of IFCs in the study area.	84
Figure 19: (A) Geological cross-section showing the typical relief and vegetation of the study area. (B) Zoomed figure showing a cross-section of an iron formation cave.	85
Figure 20: Weights adopted for each conditioning factor in the AHP analysis.	87
Figure 21: Schematic chart illustrating the methodological steps adopted in this study.	89
Figure 22: Random Forest Accuracy based on the number of trees.	93
Figure 23: Heatmap representation of the Spearman's correlation matrix for the seven research variables.	95
Figure 24: Susceptibility map of IFC occurrences, drawn using the Random Forest Classifier.	97
Figure 25: ROC Curves: (a) success rate curve using the training dataset; and (b) predicted rate curve using the validating dataset.	98
Figure 26: Performance evaluation metrics.	99
Figure 27: Variable importance expressed in terms of Mean Decrease in Gini Impurity.	100

Figure 28: Frequency Ratio values for each susceptibility class generated by Fuzzy Logic and AHP (Nola and Bacellar, 2021), and by Random Forest (this study).	101
Figura 29: 1) Localização do PNSG em um contexto regional com destaque para o polígono representativo da área de estudo. 2) Imagem de satélite da área de estudo onde se observa o relevo, a vegetação e a intervenção antrópica (solo exposto de coloração laranja). Nessa figura a letra C indica a região onde predomina a vegetação de campos rupestres ferruginosos, a letra F indica floresta estacional semidecídua, e os triângulos representam algumas quebras de relevo na área.	106
Figura 30: Arquitetura da MLP utilizada.	111
Figura 31: Representação esquemática da aplicação de Redes Neurais Artificiais neste estudo.	112
Figura 32: Variação da acurácia de treinamento e teste para diferentes números de neurônios na camada intermediária.....	115
Figura 33: Variação da acurácia de treinamento e teste para 30 reamostragens aleatórias do banco de dados de treinamento.	116
Figura 34: Mapa de suscetibilidade a ocorrência de cavidades ferríferas utilizando ANN.....	117

LISTA DE TABELAS

Tabela 1: Principais processos de gênese e evolução por erosão das cavidades ferríferas, tratados por Dutra (2013, 2015)	25
Tabela 2: Dimensões médias das cavidades ferríferas localizadas no Quadrilátero Ferrífero e seu contexto geológico. Modificado de Auler et al. (2019).....	27
Tabela 3: Legislações brasileiras em vigor. Modificado de Valentim (2016).....	29
Tabela 4: Escala de importância entre os critérios.	35
Tabela 5: Modelo de matriz de comparação em pares.	36
Tabela 6: Valores do Índice Randômico (RI).	37
Tabela 7: Resumo das justificativas para a escolha de cada variável.....	55
Table 8: Control points from the fuzzy relevance function to the criteria.	71
Table 9: Scale of importance between the criteria (Saaty 1996, 2004).....	72
Table 10: Matrix for pair comparison, criteria weights, and consistency ratio.	74
Table 11: Random consistency index (RI).	75
Table 12: Comparison between susceptibility classes and caves.	76
Table 13: Main reasons for choosing the conditioning factors adopted in this study.	88
Table 14: Database and geoprocessing environments of the conditioning factors.....	91
Tabela 15: Métricas de avaliação de desempenho.	116

LISTA DE SIGLAS

AHP	<i>Analytic Hierarchy Process</i>
ALOS	<i>Advanced Land Observing Satellite</i>
ANM	Agência Nacional de Mineração
ANN	<i>Artificial Neural Network</i>
ANP	<i>Analytic Network Process</i>
APA	Área de Proteção Ambiental
ASF	<i>Alaska Satellite Facility</i>
AUC	<i>Area Under Curve</i>
BIF	<i>Banded Iron Formation</i>
CANIE	Cadastro Nacional de Informações Espeleológicas
CAPES	Coordenação de Aperfeiçoamento de Pessoal de Nível Superior
CECAV	Centro Nacional de Pesquisa e Conservação de Cavernas
CI	<i>Consistency Index</i>
CMP	<i>Carajás Mineral Province</i>
CNC	Cadastro Nacional de Cavernas
CNPq	Conselho Nacional de Desenvolvimento Científico e Tecnológico
CNS	Cavidade Natural Subterrânea
CODEMIG	Companhia de Desenvolvimento Econômico de Minas Gerais
CR	<i>Consistency Ratio</i>
CS	<i>Caves Susceptibility</i>
DEM	<i>Digital Elevation Model</i>
DM	<i>Data Mining – DM</i>
DT	<i>Decision Tree</i>

EIA	Estudo de Impacto Ambiental
ESA	<i>European Space Agency</i>
ESRI	<i>Environmental Systems Research Institute</i>
FAPEMIG	Fundação de Amparo à Pesquisa do Estado de Minas Gerais
FN	<i>False Negative</i>
FP	<i>False Positive</i>
FR	<i>Frequency Ratio</i>
GIS	<i>Geographic Information System</i>
HAND	<i>Height Above Nearest Drainage</i>
IEF	Instituto Estadual de Florestas
IFC	<i>Iron Formation Cave</i>
IJS	<i>International Journal of Speleology</i>
IOR	<i>Iron Oxide Ratio</i>
IQ	<i>Iron Quadrangle</i>
KDD	<i>Knowledge Discovery in Database</i>
LASER	<i>Light Amplification by Stimulated Emission of Radiation</i>
LI	<i>Lineaments Intensity</i>
LVQ	<i>Learning Vector Quantization</i>
MACBETH	<i>Measuring Attractiveness by a Categorical Based Evaluation Technique</i>
MCA	<i>Multi-criteria Analysis</i>
MDT	Modelo Digital de Terreno
ML	<i>Machine Learning</i>
MLP	<i>Multilayer Perceptron</i>
MSI	<i>Multi-Spectral Instrument</i>
NDVI	<i>Normalized Difference Vegetation Index</i>
NFR	<i>Normalised Frequency Ratio</i>

NIR	<i>Near infrared</i>
NUGEO	Núcleo de Geotecnia da Escola de Minas
PALSAR	<i>Phased Array type L-band Synthetic Aperture Radar</i>
PLC	<i>Plan Curvature</i>
PRC	<i>Profile Curvature</i>
PROMETHEE	<i>Preference Ranking Organization METHod for Enrichment of Evaluations</i>
REIS	<i>RapidEye Earth Imaging System</i>
ReLu	<i>Rectified Linear Unit</i>
RF	<i>Random Forest</i>
RI	<i>Random Consistency Index</i>
RMBH	Região Metropolitana de Belo Horizonte
ROC	<i>Receiver Operating Characteristic</i>
SGNP	<i>Serra do Gandarela National Park</i>
SP	<i>Slope Gradient</i>
TIN	<i>Triangulated Irregular Network</i>
TN	<i>True Negative</i>
TOA	<i>Top Atmosphere Reflectance</i>
TOPSIS	<i>Technique for Order of Preference by Similarity to Ideal Solution</i>
TP	<i>True Positive</i>
UFOP	Universidade Federal de Ouro Preto
UTM	Universal Transversa de Mercator
WGS	<i>World Geodetic System</i>
WLC	<i>Weighted Linear Combination</i>

SUMÁRIO

CAPÍTULO 1	16
INTRODUÇÃO	16
1.1 JUSTIFICATIVA	18
1.2 OBJETIVOS	19
1.3 ORGANIZAÇÃO DO DOCUMENTO	19
CAPÍTULO 2	20
REVISÃO BIBLIOGRÁFICA	20
2.1 CAVIDADES NATURAIS SUBTERÂNEAS EM LITOTIPOS FERRUGIN	20
2.1.1 Gênese das Cavidades Ferríferas	22
2.1.2 Características das cavidades ferríferas	26
2.2 LEGISLAÇÃO ESPELEOLÓGICA BRASILEIRA E O CONTEXTO MIN	29
2.3 MÉTODOS DE ANÁLISE E PREDIÇÃO	33
2.3.1 Processo Hierárquico Analítico	33
2.3.2 Aprendizado de Máquina	38
CAPÍTULO 3	52
MATERIAIS E MÉTODOS	52
3.1 PRIMEIRA ETAPA	52
3.2 SEGUNDA ETAPA	55
3.3 TERCEIRA ETAPA	57
CAPÍTULO 4	59
MULTI-CRITERIA ANALYSIS FOR MAPPING SUSCEPTIBILITY TO IRON FORMATION CAVES DEVELOPMENT IN THE GANDARELA MOUNTAIN RANGE (MG), SOUTHEAST BRAZIL	59
4.1 INTRODUCTION	60
4.2 STUDY AREA	62
4.3 MATERIALS AND METHODS	64
4.3.1 Criteria influencing the development of iron formation caves	64
4.3.2 Data processing	68
4.3.3 Standardisation of criteria using fuzzy sets theory	70
4.3.4 Analytical Hierarchy Process (AHP)	72
4.3.5 Iron Formation Caves Susceptibility	72
4.3.6 Validation	73

4.4	RESULTS.....	73
4.5	CONCLUSIONS.....	78
	CAPÍTULO 5.....	80
	MAPPING THE SUSCEPTIBILITY TO THE OCCURRENCE OF IRON FORMATION CAVES USING THE RANDOM FOREST CLASSIFIER	80
5.1	INTRODUCTION.....	81
5.1.1	The study area	83
5.1.2	Genetic Factors.....	87
5.2	MATERIALS AND METHODS	89
5.2.1	IFC Inventory	90
5.2.2	Data Input.....	90
5.2.3	Random Forest Classifier	91
5.2.4	Training and Prediction	93
5.2.5	Validation and Evaluation	94
5.3	RESULTS AND DISCUSSIONS	94
5.4	CONCLUSIONS	101
	CAPÍTULO 6.....	103
	REDES NEURAIAS ARTIFICIAIS APLICADAS AO MAPEAMENTO DA SUCETIBILIDADE À OCORRÊNCIA DE CAVIDADES FERRÍFERAS....	103
6.1	INTRODUÇÃO.....	104
6.1.1	Área de Estudo	105
6.1.2	Cavidades Ferríferas.....	107
6.2	ESTUDOS ANTERIORES	108
6.3	MATERIAIS E MÉTODOS	110
6.3.1	Ajuste dos parâmetros e treinamento	112
6.3.2	Avaliação do desempenho da ANN	114
6.4	RESULTADOS	114
6.4.1	Análise de sensibilidade dos modelos	114
6.4.2	Suscetibilidade à ocorrência de cavidades ferríferas.....	117
6.5	DISCUSSÃO DOS RESULTADOS	118
6.6	CONCLUSÃO.....	120
	CAPÍTULO 7.....	121
	CONCLUSÃO.....	121
	REFERÊNCIAS BIBLIOGRÁFICAS	123

CAPÍTULO 1

INTRODUÇÃO

Cavidades naturais são cenários para espeleoturismo, prática de esportes radicais, compondo tela para pinturas rupestres e guardando acervos arqueológicos e paleontológicos de grande importância. Apesar das cavidades desenvolvidas em rochas carbonáticas serem as mais conhecidas num contexto mundial e objeto de estudos há muitos anos, cavidades hospedadas em outros tipos de rochas também podem ser desenvolvidas e são encontradas com frequência em todo o mundo. Como exemplos brasileiros destes tipos de cavidades (não-carbonáticas) tem-se a Gruta dos Ecos (Cocalzinho/GO), inserida em micaxistos (Cadamuro, 2001), cavernas quartzíticas em Carrancas e Luminárias (Minas Gerais) (Neto e Baptista Filho, 1997) ou na Chapada Diamantina (Bahia) (Laureano, 1998), e em rochas graníticas, como a Gruta do Riacho Subterrâneo (Itu/SP) (Rocha, 2015).

Ambientes ferruginosos também podem desenvolver cavidades e os primeiros registros desse pseudocarste datam do século XIX (Souza e Carmo, 2015). Apesar do grande aumento das pesquisas sobre cavidades ferríferas na última década, ainda há muito o que se desenvolver neste campo de atuação e uma das motivações é a estreita relação entre elas e os depósitos minerais de ferro.

Depósitos minerais de ferro são de grande importância para a economia brasileira e para empresas mineradoras que abastecem o consumo crescente deste bem mineral, e as cavidades naturais são bastante comuns neste ambiente, visto as suas disposições rochosas, a composição mineralógica e o processo de intemperismo químico/físico atuante sobre esses depósitos. Segundo o Decreto 6640/08 da legislação brasileira (Brasil, 2008), as cavidades naturais subterrâneas com relevância biológica, ou física, ou ecológica, ou histórica-cultural, devem ser preservadas para não sofrerem impactos negativos irreversíveis, preservando também uma área de influência. Por conseguinte, a ocorrência de uma ou mais cavidades naturais subterrâneas pode afetar sensivelmente o planejamento e o avanço da lavra, imobilizando por vezes reservas expressivas de minério.

Auler e Piló (2015) apontam que em áreas mineráveis, a densidade das cavidades dentro do minério de ferro tende a ser alta, com média de aproximadamente 0,15 cavernas/hectare e, levando em consideração que o comprimento mínimo para determinação de uma cavidade é muito pequeno, de 5 m, qualquer área com afloramentos rochosos pode conter uma feição deste tipo.

Atualmente a identificação de cavidades é realizada por meio de prospecção espeleológica, que pode ser muito complexa e custosa. Esta técnica precede outros estudos mais específicos, como a caracterização do meio biótico e abiótico da feição.

Em vista disso, os métodos para predição da ocorrência de cavidades ferríferas oferecem uma previsão rápida, remota e de baixo custo de áreas favoráveis à prospecção espeleológica, otimizando o trabalho de campo, diminuindo o tempo e o investimento financeiro em prospecção, e, beneficiando também o licenciamento ambiental por meio do prognóstico de novas áreas para compensação espeleológica.

Existem diversos métodos preditivos, podendo-se destacar a Análise Multicritério (*Multi-criteria Analysis* – MCA) utilizando o Processo Hierárquico Analítico (*Analytic Hierarchy Process* – AHP). O AHP é um dos métodos de MCA mais amplamente aplicável e se baseia em procedimentos de comparação entre as variáveis, para pontuar a sua importância e modelar um problema complexo (Dodgson et al., 2009). Outro método de crescente aplicação, é o Aprendizado de Máquina (*Machine Learning* – ML). O Aprendizado de Máquina busca explorar um conjunto de informações a fim de buscar padrões e relações entre as variáveis que representam o problema. Dentre as técnicas utilizadas, tem-se: Floresta Aleatória (*Random Forest* - RF) e Redes Neurais Artificiais (*Artificial Neural Network* - ANN).

Portanto, buscou-se realizar uma pesquisa que contribuísse para a compreensão da formação e ocorrência das cavidades ferríferas à luz de métodos preditivos amplamente utilizados em diversas áreas do conhecimento, culminando na avaliação do potencial espeleológico para estas feições.

1.1 JUSTIFICATIVA

O Brasil possui um enorme potencial espeleológico, e o número de cavidades ainda não conhecidas em litologias ferruginosas pode chegar a 10.000, apesar de somente cerca de 20% dessa estimativa ser conhecida (Piló e Auler, 2009). A identificação dessas cavidades também é importante para a arqueologia, a paleontologia e a biologia (Ganem, 2009; Pereira e Souza, 2009) devido à presença de vestígios arqueológicos e paleontológicos que ajudam a compreender o processo de ocupação desses ambientes (Baeta e Piló, 2005; Carmo et al., 2011) e à biodiversidade espeleológica, em especial aos animais exclusivamente cavernícolas, os troglóbios (Iniesta e Ferreira, 2013).

O processo de identificação e caracterização das cavidades envolvem a exposição dos profissionais ao risco de acidentes em trabalhos de campo, consomem muito tempo de pesquisa, e aumentam os custos financeiros que envolvem as prospecções espeleológicas. De fato, a prospecção espeleológica pode ser complexa e custosa a depender de vários fatores, como: a segurança dos profissionais, o difícil acesso à área, o tempo gasto para concluir o trabalho, entre outros. Segundo Auler e Piló (2015), uma equipe de espeleologia de 3 pessoas faz o levantamento de 2 a 3 cavidades por dia para cada uma das linhas de pesquisa (geológica, bioespeleológica e morfológica). Os autores pontuam que o esforço total necessário para um estudo completo de avaliação de cavidades é significativo, considerando o tempo de laboratório e de produção de relatórios, e envolve uma equipe de não menos que 25 pessoas especializadas em um período de aproximadamente 1,5 ano.

Além disso, o cumprimento das responsabilidades jurídicas junto aos órgãos ambientais requer, para o atendimento da compensação espeleológica, a aquisição de novas áreas nas regiões de minério de ferro, fora dos limites de exploração mineral.

Diante do exposto, o trabalho em questão possui aplicação prática e se justifica pela contribuição em um assunto até então pouco explorado pela comunidade acadêmica - apesar de haver diversos trabalhos de destaque na área de espeleologia de cavidades ferríferas relacionados à gênese, ecossistema cavernícola, grau de relevância e área de influência, motivados pelas necessidades empresariais.

1.2 OBJETIVOS

Esta pesquisa tem como objetivo principal avaliar a aplicação de métodos preditivos (AHP, RF e ANN) no mapeamento da suscetibilidade à ocorrência de cavidades naturais hospedadas em litotipos ferruginosos, a partir de um banco de dados de caracterização do meio físico, em busca de áreas-alvo para o mapeamento espeleológico.

Para alcançar o objetivo principal, são definidos alguns objetivos específicos como:

- Estudar e compreender os processos de formação das cavidades ferríferas para escolha das variáveis condicionantes na área de estudo;
- Avaliar a viabilidade de utilização das técnicas empregadas no mapeamento de cavidades ferríferas;
- Avaliar as vantagens e desvantagens do procedimento utilizado neste estudo.

1.3 ORGANIZAÇÃO DO DOCUMENTO

Este documento foi estruturado em formato de artigos. O capítulo 1 apresenta a introdução e justificativa, bem como os objetivos que geraram as motivações necessárias a realização deste estudo. O capítulo 2 apresenta uma breve revisão bibliográfica sobre os temas abordados. O capítulo 3 refere-se aos materiais e métodos que envolveram a execução da tese a partir de um roteiro metodológico aplicado para alcançar os objetivos. O capítulo 4 apresenta o primeiro artigo, publicado, destinado a avaliar a aplicação do método de processo de análise hierárquica e lógica fuzzy para mapeamento da suscetibilidade à ocorrência de cavidades ferríferas. O capítulo 5 apresenta o segundo artigo, submetido, destinado a avaliar a aplicação de Floresta Aleatória no mapeamento da suscetibilidade à ocorrência de cavidades ferríferas. O capítulo 6 apresenta o terceiro artigo, ainda não submetido, destinado a fazer uma comparação entre as técnicas utilizadas nos estudos anteriores e os resultados obtidos na aplicação de Redes Neurais Artificiais. E por último, o capítulo 7 traz a conclusão geral desta pesquisa e as recomendações para trabalhos futuros.

CAPÍTULO 2

REVISÃO BIBLIOGRÁFICA

2.1 CAVIDADES NATURAIS SUBTERRÂNEAS EM LITOTIPOS FERRUGINOSOS

Uma cavidade natural subterrânea (CNS), de acordo com o Decreto 6640/08, “ é todo e qualquer espaço subterrâneo acessível pelo ser humano, com ou sem abertura identificada...” (Brasil, 2008). Este espaço, também chamado de caverna, gruta, toca e buraco, desde a pré-história possui uma relação direta com seres vivos de diferentes espécies. Trata-se de um lugar que serve de moradia e abrigo para animais, onde são encontrados diversos vestígios arqueológicos, guarda registros fossilíferos, e serve de habitat para desenvolvimento de fauna e flora cavernícolas. Assim, tão importante quanto a feição geomorfológica da cavidade também é seu conteúdo mineral, hídrico, biótico, arqueológico, paleontológico e geológico (Ganem, 2009).

As CNS se desenvolvem com maior frequência em litotipos solúveis, como calcários, cujos terrenos são definidos como cársticos, em referência à região carbonática situada na fronteira entre a Itália e a Eslovênia, considerada o Carste Clássico (Piló e Auler, 2013). Em todo o mundo, as CNS também são encontradas em quartzitos e arenitos, em formações ferríferas, além de litotipos como xistos, granitos e gnaisses (White et al., 1966; Aubrecht et al., 2008; Karmann et al., 2001; Oberender et al., 2014; Welianje et al., 2010).

No Brasil, os primeiros relatos científicos sobre cavidades são da metade do século XIX, data dos trabalhos do naturalista Peter Lund acerca das cavernas encontradas na região de Lagoa Santa (MG), e do seu patrimônio paleontológico (Figueiredo, 2011). Atualmente, de acordo com o Cadastro Nacional de Cavernas (CNC, 2021), das 8172 CNS cadastradas até o momento, 70,1% correspondem às cavernas inseridas em calcários, dolomitos ou mármore, 12,5 % em rochas como arenitos, conglomerados e argilitos, 11,8% em rochas metassedimentares (quartzitos e formações ferríferas), 0,7% em depósitos supergênicos (bauxita e canga) e, os 4,9% restantes em outros litotipos como granitos, basaltos, gnaisses, xistos, entre outros.

Com base nessas informações, observa-se que as cavidades ferríferas são pouco expressivas quando comparadas às de ambiente cárstico. Mesmo assim, as cavidades naturais subterrâneas em litotipos ferruginosos são objeto de estudo há décadas, como no trabalho pioneiro de Simmons (1963) na região do Quadrilátero Ferrífero (*Iron Quadrangle* - IQ). E, atualmente, as prospecções espeleológicas nestes litotipos se expandem, devido à exigência de realização dos Estudos de Impacto Ambiental (EIA) para instalação de empreendimentos minerários e exploração de minério de ferro.

As cavidades naturais subterrâneas desenvolvidas em litotipos ferruginosos são encontradas no sistema geomorfológico chamado Geossistema Ferruginoso e, neste estudo, serão nomeadas como cavidades ferríferas (*Iron Formation Caves* – IFCs) (Figura 1). Para Christofolletti (1980), um sistema em geomorfologia é “o conjunto dos elementos e das relações entre si e seus atributos”. Sendo assim, geossistemas ferruginosos se referem às unidades territoriais naturais, resultante da combinação dinâmica entre os fatores fisiográficos, biológicos e antrópicos e constituídas por rochas como itabiritos, cangas, jaspilitos entre outras. (Souza e Carmo, 2015).

Com efeito, destaca-se a importância de se compreender a evolução dos conhecimentos em cavidades naturais localizadas em geossistemas ferruginosos visto que há muito o que se discutir e pesquisar nesta área, considerando o potencial espeleológico em ambientes ricos em ferro do território brasileiro.

Atualmente, os geossistemas ferruginosos Quadrilátero Ferrífero e Serra de Carajás/PA na Província Mineral de Carajás abrigam as mais importantes regiões espeleológicas brasileiras em se tratando de cavidades ferríferas (Piló, Coelho e Reino, 2015). No Quadrilátero Ferrífero, as cavidades ferríferas encontram-se associadas às formações ferríferas da Formação Cauê, pertencente ao Grupo Itabira, datadas do Paleoproterozoico. Esta formação é constituída por itabiritos, itabiritos dolomíticos, filitos dolomíticos, dolomitos ferruginosos e filitos (Alkmim e Marshak, 1998). As cavidades podem ainda se desenvolver em depósitos supergênicos, tanto em cangas quanto no contato entre a canga e a formação ferrífera. Em Carajás, as cavidades ferríferas estão associadas à Formação Carajás e, assim como no Quadrilátero Ferrífero, também são encontradas em cangas. Na sua grande maioria ocorrem na presença desses dois litotipos (Piló e Auler, 2009). A Formação Carajás é pertencente ao Grupo Grão-Pará, do

Neoarqueano, constituída por jaspilitos com bandas que variam entre óxidos de ferro, jaspe e chert branco, com carbonatos subordinados (Beisiegel et al., 1973).



Figura 1: Caverna ferrífera, encontrada na Serra de Carajás. Foto cedida por Georgete Dutra.

2.1.1 Gênese das Cavernas Ferríferas

As postulações relacionadas à gênese das cavernas ferríferas se iniciaram com os estudos de Simmons (1963) e até hoje são discutidas pelos pesquisadores. Todavia, uma melhor caracterização, tanto na descrição do ambiente quanto do entendimento de sua formação cavernas ferríferas, é incentivada pela comunidade espeleológica, pelas empresas mineradoras e pelos órgãos ambientais.

Simmons (1963) descreve que a gênese das cavernas ferríferas está associada aos processos de dissolução, principalmente da dolomita, mas também de quartzo e hematita. Essa dissolução pode conferir à zona alterada alta porosidade, chegando a atingir 50% do volume. Segundo Auler et al. (2019), a teoria da dissolução do cimento carbonático não é dominante na geração

de cavidades ferríferas, visto que a maioria delas se encontram no contato entre a canga e a formação ferrífera e, algumas inclusive se encontram inteiramente na formação ferrífera, o que faz a dedução de Simmons (1963) não ser totalmente aplicável.

Para Calux (2013) e Auler et al. (2014), o início da formação de uma cavidade ferríferas acontece em ambiente hipogênico, ou seja, profundo e saturado, envolvendo ação biológica e química, de dissolução da sílica e mobilização do ferro. Esse processo formaria uma espécie de câmara ou “bolha”, de vários tamanhos, encontrada em profundidade e abaixo do nível freático (Calux, 2013) (Figura 2). Um fato que corrobora com essa hipótese é a ocorrência de diversas cavidades ferríferas oclusas, em profundidade, encontradas quando é realizada uma perfuração mecânica pela empresa mineradora (Auler et al., 2019). A evolução desse contexto genético, se daria pelo soerguimento dessas regiões por meio de movimentos epirogenéticos e isostáticos, onde as cavidades ferríferas se aproximariam da superfície, enquanto se exporiam aos processos erosivos de evolução das escarpas que, conseqüentemente, conectariam esses espaços isolados e formariam as suas entradas (Figura 2). A ausência de uma rede de fluxo de água subterrânea e a falta de conectividade entre os poros, durante a fase freática, sugere que a formação dessas cavidades seja lenta, resultando em câmaras isoladas e irregulares que, na fase vadosa, se interligam por condutos estreitos e inclinados (Auler et al., 2014). Esse processo pode mascarar a morfologia original oriunda da fase freática, gerando um padrão alongado e paralelo à superfície do relevo, resultante da incorporação dos processos hidrológicos da encosta (Auler et al., 2014).

De modo geral, Calux (2013) discute a espeleogênese de cavidades ferríferas a partir de três premissas: I) As cavidades têm sua gênese e evolução condicionadas a determinados compartimentos da paisagem; II) A dinâmica evolutiva das cavidades pode ser associada à padrões morfológicos em diversas escalas (micro, macro e meso); III) O arranjo litoestrutural, as propriedades físico-químicas e a composição das rochas estão diretamente associadas à formação e evolução das cavidades.

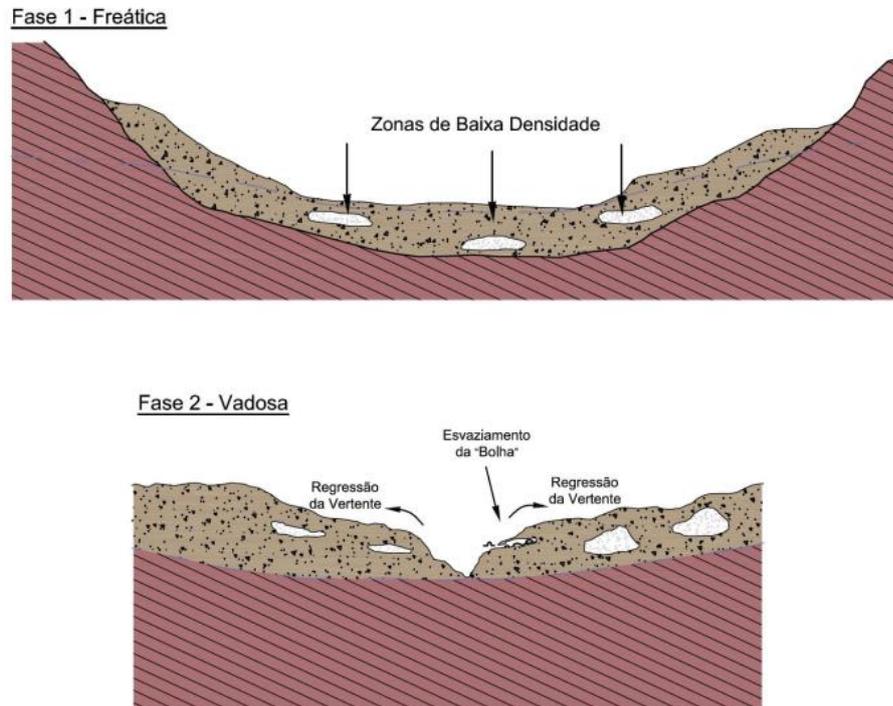


Figura 2: Modelo de gênese e evolução criado para a Caverna de Junção/MG, que ilustra a formação das cavidades em ambiente vadoso (Calux, 2013).

Além do modelo genético de hipogenia sugerido por Auler et al. (2014), Dutra (2013, 2015) discute a gênese das cavidades ferríferas a partir dos seguintes processos: erosão, lixiviação, dissolução e biogênese. Abaixo apresentam-se os conceitos e argumentos discutidos pela autora para os processos genéticos citados:

- Erosão (mecânica): O processo de erosão diferencial atua no material mais friável, geralmente localizado abaixo das cangas, que por serem materiais mais resistentes às ações da erosão, estruturam o teto das cavidades ferríferas. Esse processo pode ser dividido de acordo com a forma de ocorrência como mostra a Tabela 1
- Lixiviação: Alteração supergênica por lenta remoção do quartzo ou carbonato gerando alta porosidade no material alterado, redução do volume original deste, conseqüentemente, formação de vazios. O ferro permanece no perfil em forma de plasmas hematíticos ou goethíticos preenchendo vazios deixados pela alteração.
- Dissolução: Acontece pela dissolução por fluxos hidrotermais dos minerais constituintes da rocha e também pela migração da sílica e do ferro, em lenta

solubilização química a depender do pH da água. Forma-se um cimento hemático e/ou gibbissítico que envolve os fragmentos desagregados. Durante a evolução do processo, os fragmentos, sendo menos estáveis quimicamente do que o cimento, são lixiviados gerando porosidades sustentadas pelo cimento da dissolução.

- Biogênese: Geração ou ampliação de cavidades por paleobiotas, como tatus e preguiças gigantes, da fauna pleistocênica, em material terroso relativamente tenro. Também podem ser geradas por ação de raízes de plantas.

Tabela 1: Principais processos de gênese e evolução por erosão das cavidades ferríferas, tratados por Dutra (2013, 2015)

Em cabeceiras de drenagem	Principal processo de formação das cavidades ferríferas no contato canga/rocha. Acontece por erosão mecânica das rochas nas cabeceiras de drenagem em forma de leque, na alta e média vertente. Associada ao escoamento difuso, às nascentes sazonais situadas abaixo das cangas, gerando erosão do tipo <i>piping</i> . Assim, a canga permanece praticamente inalterada devido à sua resistência e o material mais friável vai sendo erodido progressivamente para montante. A sua evolução causa desmoronamentos laterais e do teto, principalmente em período chuvoso e, conseqüentemente, favorece à junção de cavidades próximas. O perfil dessas cavidades acompanha o contato canga/rocha e o relevo.
Remontante	Acontece quando o nível de água intercepta a superfície do terreno, formando uma nascente. A erosão avança para montante, acompanhando o nível d'água.
Por cachoeira	Um caso particular da erosão remontante. Acontece quando a água superficial em forma de cachoeira, escoar através da encosta descendo por quedas de relevo abruptos. A energia potencial da queda possui grande poder erosivo. A erosão acompanha o relevo e avança a partir da entrada erodida.
Nas margens de drenagens	Acontece quando, em período chuvoso, o fluxo d'água das drenagens provoca a erosão mecânica das margens por solapamento e, conseqüente desmoronamento do material sobrejacente. Formam abrigos de abertura longa, altura pequena e pouca profundidade.
Em bordas de lagoas	Acontece devido à erosão nas bordas de lagoas em contato direto com a linha d'água. Espera-se que as cavidades ferríferas, neste contexto, tenham maior amplitude na linha d'água das lagoas, com perfil de entrada lenticular horizontal.

Tabela 1: Principais processos de gênese e evolução por erosão das cavidades ferríferas, tratados por Dutra (2013, 2015)

Em depósitos de talus	Acontece quando fragmentos de rocha são empilhados formando cavidades e abrigos entre eles. O depósito de talus geralmente é formado pelo transporte dos fragmentos de uma área topograficamente mais alta que se depositam no sopé das encostas.
------------------------------	---

2.1.2 Características das cavidades ferríferas

As cavidades ferríferas apresentam geralmente dimensões reduzidas, possuindo uma espessura de teto diminuta, e, inclinação e perfil ascendentes acompanhando o mergulho da encosta (Dutra, 2017). Uma representação esquemática do perfil típico desse tipo de cavidade pode ser observada na Figura 3.

Como consequência da erosão diferencial, as áreas moldadas sobre litotipos ferruginosos do Quadrilátero Ferrífero se mostram mais elevadas devido à alta resistência desses materiais ao intemperismo químico e mecânico (Salgado et al., 2004). Segundo Piló et al. (2015), as cavidades ferríferas, neste geossistema, são mais encontradas no compartimento de cimeira (altas e médias vertentes). A Tabela 2 mostra as dimensões médias observadas nas cavidades ferríferas, da região do Quadrilátero Ferrífero. Nota-se que o comprimento médio de cavidades ferríferas nesta região é de 24,30 m, enquanto que a maior cavidade ferrífera já encontrada possui 372 metros de comprimento (Auler et al., 2019).

Na Serra do Gandarela, região pertencente ao Quadrilátero Ferrífero, Coelho et al. (2010) apontam que grande parte das cavidades estão posicionadas nas bordas das coberturas de canga em razão da existência de escarpas resultantes do recuo erosivo. As maiores concentrações estão entre 1.400 e 1.600 m de altitude.

De modo geral, as cavidades ferríferas se caracterizam por pequenos salões únicos de planimetria irregular e, quando maiores, por condutos muito irregulares conectados a outros condutos menores de origem tardia (Piló et al., 2015). Podem apresentar a presença de pilares, pendentes, canalículos e claraboias.

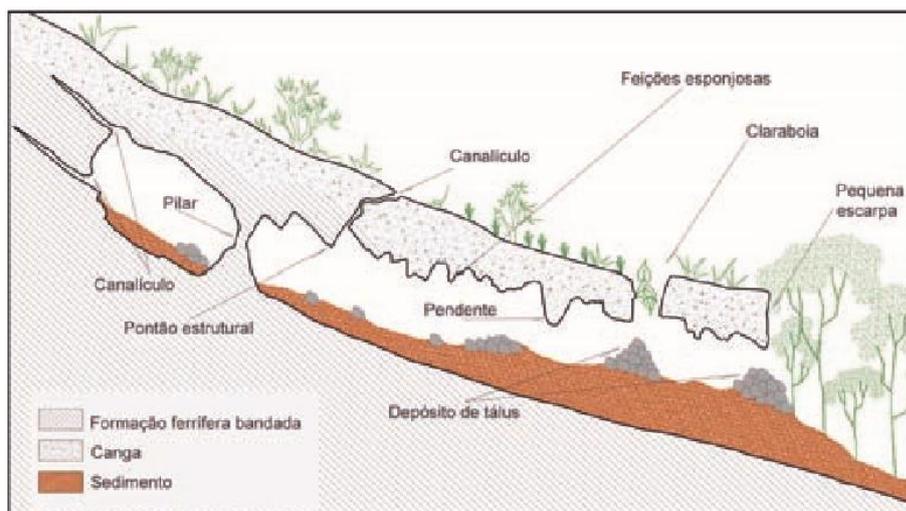


Figura 3: Perfil síntese de uma cavidade ferrífera com as principais feições morfológicas. Fonte: Piló et al., (2015).

Tabela 2: Dimensões médias das cavidades ferríferas localizadas no Quadrilátero Ferrífero e seu contexto geológico. Modificado de Auler et al. (2019).

Litotipos	Número de cavidades	Comprimento (m)	Profundidade (m)	Área (m ²)	Volume (m ³)
Canga	160	24,34	2,53	59,20	66,97
Contato entre canga e rocha	73	29,65	3,43	71,86	90,78
Formação Ferrífera Bandada (sã ou alterada)	109	20,67	3,22	55,53	78,39
Total/Média	342	24,30	2,94	60,73	75,69

Feições esponjosas ou rugosas são vistas nas paredes devido a ação constante da erosão/dissolução (Dutra, 2017). No seu interior não há atualmente fluxos de água, sendo em sua grande maioria consideradas secas (Dutra, 2013). Durante períodos chuvosos, a água pluvial pode atravessar pequenos condutos e/ou discontinuidades geológicas provocando um gotejamento significativo dentro das cavidades, o que pode resultar em fontes e canais de drenagem temporários (Piló et al., 2015).

Os espeleotemas são apresentados como coraloides que podem possuir morfologia nodular, globular, botrioidal ou similares a corais, constituídos por óxidos e hidróxidos de ferro e alumínio, fosfatos e sulfatos (Silva e Santos Junior, 2015). E também por espeleotemas do tipo crosta de cor branca ou amarela, formada pela presença de microrganismos (bactérias, líquens e musgos). Em geral os espeleotemas encontrados em cavidades ferríferas são menores do que os encontrados em ambiente carbonático.

Piló et al. (2015) discutem que as cavidades ferríferas apresentam controle estrutural, particularmente naquelas desenvolvidas em formações ferríferas. Descontinuidades podem ser observadas nas paredes e nos tetos e, devido a ruptura constante dessas descontinuidades, são bastante encontrados fragmentos de rocha/canga variando de centimétricos a métricos no piso das cavidades (Dutra, 2017). Estes fragmentos, juntamente com outros materiais mais finos, formam depósitos clásticos classificados como autogênicos, oriundos do colapso de partes do teto e das paredes (Piló, et al., 2015). Calux (2013) conclui que os principais fatores de controle estrutural das cavidades dos Quadrilátero Ferrífero são: bandamento, fraturas tectônicas e atectônicas, eixos e flancos de dobras e, o contato geológico erosivo entre a canga e a formação ferrífera bandada.

As cavidades ferríferas possuem grande relevância no cenário espeleológico por apresentarem importância arqueológica e paleontológica, além de um conjunto de características como: gênese, morfologia, espeleotemas e condições geológicas – consideradas raras e em certas esferas até únicas, abrigando formas específicas de vidas, muitas delas exclusivas desses ambientes. Elas são palco de uma biodiversidade peculiar, contendo organismos extremamente complexos, constituindo populações restritas ao ambiente subterrâneo rico em ferro, úmido e de baixa ou nenhuma luminosidade (Ferreira, 2005). Estes ambientes cavernícolas são considerados os mais ricos em espécies, principalmente troglomórficos, quando comparados às cavidades desenvolvidas em quartzito ou calcário (Souza-Silva et al., 2011), sendo destacado como um ecossistema frágil e delicado (Ribas e Carvalho, 2009).

Diante do exposto, e por outros motivos relacionados ao patrimônio ecológico, cênico, histórico-cultural e científico, a conservação das cavidades ferríferas, assim como os demais tipos de cavidades, é de extrema importância. E de forma a proteger e conservar o patrimônio espeleológico brasileiro, garantindo um desenvolvimento sustentável, o poder público se

manifesta desde a década de 80, por meio de resoluções, decretos, leis e normativas relativas a esse tema.

2.2 LEGISLAÇÃO ESPELEOLÓGICA BRASILEIRA E O CONTEXTO MINERÁRIO

A primeira vez que a espeleologia foi abordada na legislação brasileira foi na Lei Federal nº 3.924 de 26 de julho de 1961 (Brasil, 1961) que dispõe sobre “Monumentos arqueológicos e pré-históricos”, onde as cavidades foram citadas como “grutas, lapas ou abrigos sob rocha”, e foram consideradas monumentos arqueológicos e pré-históricos caso possuíssem “vestígios positivos de ocupação pelos paleoameríndio”.

Desde então, após a Constituição Brasileira em 1988, onde se institui as cavidades como bens da união, e mais significativamente a partir de 2004, com a Resolução CONAMA nº 347, vários mecanismos e regulamentações sobre o patrimônio espeleológico foram estabelecidos (Tabela 3).

Tabela 3: Legislações brasileiras em vigor. Modificado de Valentim (2016).

Ano	Esfera	Marco Regulatório	Síntese dos dispositivos
1961	Federal	Lei nº 3.924	Dispõe sobre Monumentos Arqueológicos e Pré-históricos.
1986	Federal	Resolução CONAMA nº009	Cria Comissão Especial para assuntos relativos à preservação do Patrimônio Espeleológico Nacional (PEN).
1988	Federal	Constituição Federal, Art. 20, inciso X	Define as cavidades como Bens da União.
1990	Federal	Portaria IBAMA nº 887	Promove a realização de diagnóstico das cavidades no Brasil, limita o uso das cavidades e cria a obrigatoriedade de EIA/RIMA para atividades potencialmente lesivas as cavidades.
1990	Federal	Decreto 99.556	Dispõe sobre o Patrimônio Espeleológico e delega ao IBAMA preservar, conservar, fiscalizar e controlar o uso das cavidades, bem como fomentar levantamentos e estudos necessários para a conservação integral de todas as cavidades e de sua área de influência.
2004	Federal	Resolução CONAMA nº 347	Estabelece uma área no entorno das cavidades como necessária para garantir o seu equilíbrio ecológico e manutenção da sua integridade física (área de influência), delega ao IBAMA a definição da área de influência e para tal poderão ser solicitados ao empreendedor estudos específicos. E admite a construção e instalação de atividades nocivas ao PEN, condicionadas ao licenciamento ambiental para as cavidades não relevantes.

Tabela 3: Legislações brasileiras em vigor. Modificado de Valentim (2016).

Ano	Esfera	Marco Regulatório	Síntese dos dispositivos
2008	Federal	Decreto 6.640	Dá nova redação ao Decreto 99.556 e possibilita a intervenção em cavidades conforme seu grau de relevância, estabelece as formas de compensação e os estudos necessários para a classificação da relevância. Delega ao órgão licenciador, no âmbito do processo de licenciamento a responsabilidades sobre o tema.
2009	Federal	Instrução Normativa MMA nº 02	Estabelece a metodologia para a classificação dos graus de relevância e a definição de conceitos alguns termos tratados no Decreto 6.640/08.
2009	Federal	Portaria MMA nº 358	Institui o Programa Nacional de Conservação do Patrimônio Espeleológico, que tem como objetivo desenvolver estratégia nacional de conservação e uso sustentável do PEN.
2010	Federal	Resolução CONAMA nº 428	Dá nova redação a Resolução CONAMA nº 374/04 dizendo que todas as cavidades são relevantes com diferentes graus conforme metodologia da IN MMA nº02/09 repassando a responsabilidade e anuência das cavidades para o órgão licenciador competente.
2014	Federal	Termo de Referência CECAV	Orientações Básicas à realização de estudos espeleológicos sobre a área de influência do Patrimônio Espeleológico.
2014	Estadual	Instrução de Serviço SEMAD MG nº 03	Trata sobre procedimentos administrativos a serem realizados pelas SUPRAM's quanto aos empreendimentos considerados efetiva ou potencialmente poluidores ao patrimônio espeleológico.
2016	Estadual	Decreto MG nº 47.041	Trata dos critérios para a compensação e indenização dos impactos e danos causados em cavidades.
2016	Federal	Termo de Referência CECAV	Orientações Básicas à realização de estudos ambientais sobre Sismografia Aplicada ao Patrimônio Espeleológico.
2017	Federal	Instrução Normativa MMA nº 01	Estabelece procedimentos para definição de outras formas de compensação ao impacto negativo irreversível em cavidades natural subterrânea com grau de relevância alto.
2017	Federal	Instrução Normativa MMA nº 02	Revoga a IN nº02 de 2009. Estabelece metodologia para classificação do grau de relevância das cavidades naturais subterrâneas.
2017	Federal	Instrução Normativa MMA nº 04	Acrescenta o art. 5-A à Instrução Normativa nº 1, de 24 de janeiro de 2017 que estabelece procedimentos para definição de outras formas de compensação ao impacto negativo irreversível em cavidade natural subterrânea com grau de relevância alto.

Como visto, as cavidades possuem um regime de uso e proteção próprio, e é louvável o desejo de preservação ambiental como premissa constitucional. Entretanto, Ferraz (2013) destaca que

conservar é um princípio do desenvolvimento sustentável desde que a conservação não amplie o bloqueio do uso dos recursos ambientais, recursos esses necessários para o crescimento econômico brasileiro. Assim, deve-se haver uma equidade na gestão dos recursos ambientais, nesse caso, na gestão das cavidades, definindo um sistema seguro e coerente para o uso racional desses recursos de acordo com a Política Nacional de Meio Ambiente e todos os demais instrumentos (Ferraz, 2013).

Como as cavidades ferríferas estão intimamente relacionadas às áreas de mineração de ferro e são protegidas pela legislação supracitada, a implantação de empreendimentos que podem causar poluição ou degradação em cavidades e/ou na sua área de influência é dependente de licenciamento ambiental em órgão competente. Para ser licenciado, os empreendimentos minerários precisam elaborar estudos ambientais com o objetivo de analisar uma série de critérios estabelecidos principalmente pelo Decreto 6.640 de 7 de novembro de 2008 (Brasil, 2008).

A análise desses critérios, citados no referido decreto, serve de base para a classificação das cavidades em graus de relevância, de acordo com as seguintes classes: máximo, alto, médio a baixo – determinadas “pela análise de atributos ecológicos, biológicos, geológicos, hidrológicos, paleontológicos, cênicos, histórico-culturais e socioeconômicos, avaliados sob enfoque regional e local” (Brasil, 2008).

Quando as cavidades são classificadas como de baixa relevância, ela pode ser removida e não são necessárias medidas compensatórias (Brasil, 2008). Quando classificadas como de média relevância, elas podem ter impactos negativos, sem condicionantes, e podem ser removidas desde que aconteça a compensação espeleológica por meio de medidas que contribuam para a conservação e uso adequado do patrimônio espeleológico (Brasil, 2008).

Entretanto, devido ao cenário altamente restritivo produzido pelos critérios de classificação dos graus de relevância, principalmente quando se trata de cavidades ferríferas, a maioria das cavidades acabam se enquadrando na classe de máxima e alta relevância (Auler e Piló, 2015).

Assim, ao ser considerada pertencente à classe alta, a cavidade pode ser impactada negativamente com condicionantes, ou, ser removida desde que duas outras cavidades de alta relevância, preferencialmente dentro da mesma unidade geológica e contendo características

semelhantes, estejam permanentemente protegidas (Brasil, 2008). Se considerada de máxima relevância, a cavidade deve ser protegida assim como toda a área num raio de 250 m, identificada como área de influência (Brasil, 2008). Nesta circunstância, ou a cavidade não pode ser objeto de impactos negativos irreversíveis, ou cria-se uma busca por aquisição de novas áreas onde podem haver cavidades de alta relevância para a compensação.

Piló et al. (2015b) expõem uma lista de dificuldades observadas no processo de licenciamento ambiental, quando se trata da compensação espeleológica para cavidades ferríferas de alta relevância, a saber:

- Rigidez locacional da mineração;
- Ausência de cavidades para compensação na área do empreendimento, principalmente em áreas de elevada densidade de processos minerários, como no Quadrilátero Ferrífero;
- Falta de áreas adjacentes à mineração por estarem requeridas por terceiros juntos à ANM;
- Modificação da viabilidade de jazida, que está condicionada à exploração integral da ocorrência mineral;
- Eventuais alterações no planejamento da lavra inicialmente aprovado no Plano de Aproveitamento Econômico (PEA) entregue à ANM para fins de licenciamento, culminando em impedimentos legais;
- Mudança no planejamento mineiro para garantir a preservação de cavidades testemunhos, o que pode gerar outros impactos ambientais.

Havendo impactos negativos em cavidades de alta relevância e, não existindo na área do empreendimento outras cavidades que possam ser preservadas sob a forma de cavidades testemunho, o caminho dado pela legislação é a compensação excepcional, determinada pelo órgão licenciador (Piló et al., 2015b), o que inclui, por exemplo, o fornecimento de apoio financeiro ou a criação/manutenção de uma unidade de conservação espeleológica (Auler e Piló, 2015).

Portanto, o atual panorama brasileiro leva ao seguinte dilema: a exploração total dos recursos minerais ferríferos levará à perda de cavidades ferríferas, e a preservação de cavidades ferríferas provavelmente restringirá o desenvolvimento da mineração. E este dilema possui

consequências econômicas e sociais, como reduções significativas na área explorável, que pode chegar a 50% da área lavrável inicialmente planejada, além perdas financeiras com os atrasos nos estudos espeleológicos (Auler e Piló, 2015).

Com todas essas adversidades e com o aumento do número de cavidades ferríferas prospectadas em áreas de mineração de ferro, o desafio do governo, dos órgãos ambientais e dos empreendedores minerários é preservar as cavidades sem prejudicar o desenvolvimento econômico advindo da exploração de minério de ferro. E, segundo Auler e Piló (2015), esse cenário apresenta atualmente uma das restrições mais sérias ao desenvolvimento mineral brasileiro. Lembrando que todas essas exigências legais também se aplicam aos projetos de reservatórios, linhas de eletricidade, empreendimentos urbanos, estradas e ferrovias.

Dentre todas as possibilidades de aporte científico para solucionar esse dilema, um deles é a determinação do potencial espeleológico de uma região, que irá direcionar a prospecção de cavidades e aumentará as chances de se encontrar áreas para compensação espeleológica próximas aos empreendimentos minerários. Para isso, podem ser empregadas as técnicas de Análise Multicritério, por meio de AHP e *Machine Learning* por meio de Floresta Aleatória e Redes Neurais Artificiais.

2.3 MÉTODOS DE ANÁLISE E PREDIÇÃO

2.3.1 Processo Hierárquico Analítico

A Análise Multicritério é um conjunto de técnicas para auxiliar e apoiar a tomada de decisões. Para isso, é necessário traçar um objetivo a ser alcançado e, a partir de critérios que são analisados simultaneamente, são encontradas as alternativas para a solução do problema (Dodgson et al., 2009; Ishizaka e Nemery, 2013). A abordagem da Análise Multicritério requer o exercício do julgamento, para realizar as ponderações relativas entre todos os critérios, e apenas se diferem no modo em que combinam os dados (Gimenes, 2013).

Assim, existem diversas técnicas que podem dar apoio às decisões frente a problemas como de escolha, classificação, hierarquizar opções, de descrição do problema fazendo uma distinção entre as possibilidades aceitáveis ou não, por exemplo (Dodgson et al., 2009). Entre as técnicas pode-se citar:

- AHP (*Analytic Hierarchy Process*), proposto por Saaty (1977);
- TOPSIS (*Technique for Order of Preference by Similarity to Ideal Solution*), desenvolvido por Hwang e Yoon (1981);
- PROMETHEE (*Preference Ranking Organization METHod for Enrichment of Evaluations*), proposto por Brans (1982);
- MACBETH (*Measuring Attractiveness by a Categorical Based Evaluation Technique*), desenvolvido por Bana e Costa e Vasnick (1994);
- ANP (*Analytic Network Process*), proposto por Saaty (1996).

Dentre todas essas técnicas, o Processo Hierárquico Analítico (AHP) foi a escolhida para ser aplicada nesta pesquisa.

O AHP busca solucionar o problema dividindo-o em subproblemas (Ishizaka e Nemery, 2013). Essa divisão é feita inicialmente pela estruturação do problema seguida da enumeração das prioridades por meio de comparações entre pares. Assim, o problema é estruturado como mostra a Figura 4, tendo como primeiro nível o objetivo da decisão. O segundo nível representa os critérios e o terceiro nível representa as alternativas. Caso o problema seja ainda mais complexo, outros níveis de subcritérios podem ser adicionados.

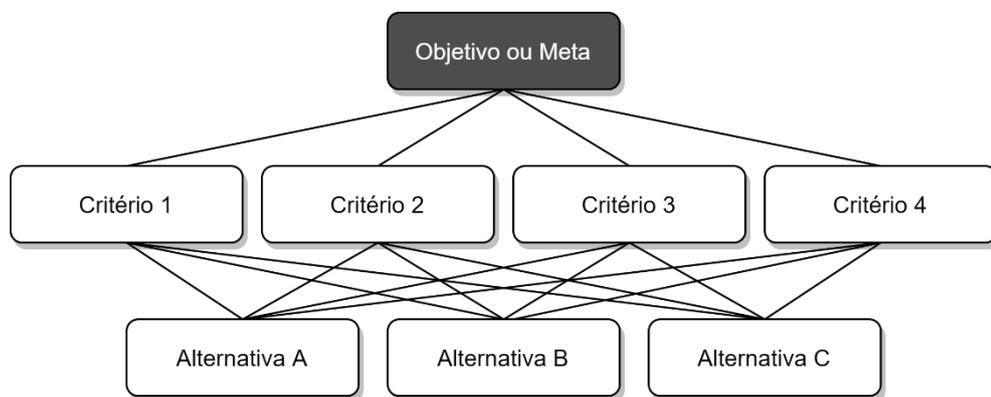


Figura 4: Estrutura hierárquica do processo de AHP.

Para classificar as alternativas, após a estruturação do problema, as prioridades são calculadas com base nas comparações de pares que podem ser feitas por um julgamento numérico ou verbal; então é feita a verificação de consistência e a análise de sensibilidade (Saaty, 2004; Dodgson et al., 2009; Tzeng e Huang, 2011). Ishizaka e Nemery (2013) sugerem que essas duas últimas etapas são opcionais, contudo, recomenda-se as suas realizações a vista de confirmar a

robustez dos resultados. A aplicação do método AHP pode ser dividida em quatro etapas (Saaty, 2001; Tzeng e Hung, 2011):

- a. **Criação de um sistema hierárquico:** nessa etapa acontece a decomposição do problema em elementos inter-relacionados na qual são elencados todos os critérios ou variáveis que compõem o problema. De uma forma geral, deve-se prever todos os caminhos pelos quais se alcança o objetivo ou a meta idealizada. No caso da aplicação proposta neste trabalho, são selecionadas as variáveis que estão associadas à gênese das cavidades ferríferas e suas principais características.

- b. **Criação da matriz de comparação em pares entre os critérios:** essa etapa é definida pela organização dos critérios em uma matriz com base na atribuição dos valores da escala linear de importância que varia de 1 a 9 (Tabela 4), sugerida por Saaty (2001). A importância de cada um dos critérios em relação aos demais critérios é avaliada e disposta nesta matriz de decisão, onde a diagonal principal é sempre igual a 1, por se tratar de comparação entre os mesmos critérios, e os dois lados da matriz são correspondentes (Tabela 5).

Tabela 4: Escala de importância entre os critérios.

Intensidade da importância	Definição	Explicação
1	Igual importância	Os dois critérios contribuem igualmente para o objetivo
3	Moderada importância	A experiência favorece um critério em relação ao outro
5	Forte importância	A experiência favorece fortemente um critério em relação ao outro
7	Muito forte importância	Um critério é muito fortemente favorecido em relação ao outro e pode ser demonstrado na prática
9	Extremamente importante	A experiência favorece um critério em relação ao outro, sendo a mais alta ordem de afirmação possível
2,4,6,8	Valores intermediários entre dois julgamentos adjacentes	Quando é necessária uma condição de compromisso

Tabela 5: Modelo de matriz de comparação em pares.

	Critério 1	Critério 2	Critério n
Critério 1	1	$a_{1,2}$	$a_{n,1}$
Critério 2	$a_{2,1} = 1/a_{1,2}$	1	$a_{2,n}$
Critério n	$a_{1,n} = 1/a_{n,1}$	$a_{n,2} = 1/a_{2,n}$	1

- c. **Sumarização dos julgamentos dos critérios:** após a etapa de comparação entre pares é determinado o peso relativo das alternativas em relação aos critérios, a partir do cálculo do autovetor normalizado desta matriz. Esta operação origina o vetor de pesos dos critérios frente às alternativas. Na sequência, para que a matriz de comparação seja consistente, é necessário que seja calculado a razão de consistência da matriz (*Consistency Ratio* - CR) e este ser comparado ao índice randômico (Tabela 6) apresentado por Saaty (2001). O mesmo autor sugere que a razão de consistência deve ter um valor limite de 0,1. Caso esse valor seja maior, os julgamentos devem ser revisados.

A razão de consistência (CR) que detecta contradições na determinação da importância relativa. O valor de CR (Equação 1) é baseado no índice de consistência (*Consistency Index* - CI) (Equação 2) e no índice randômico (*Random Consistency Index* - RI) (Tabela 6).

$$CR = CI / RI \quad (\text{Eq. 1})$$

e,

$$CI: \lambda_{\max} - n / n-1 \quad (\text{Eq. 2})$$

onde,

λ_{\max} : número principal de eigen;

n: número total de critérios;

RI: *Random Consistency Index* para cada valor de n.

Tabela 6: Valores do Índice Randômico (RI).

N	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15
RI	0	0	0,58	0,90	1,12	1,24	1,32	1,41	1,45	1,49	1,51	1,53	1,56	1,57	1,59

d. Agregação dos pesos e decisão final: nessa etapa acontece a agregação dos pesos relativos para formulação da melhor alternativa de modo a avaliar o resultado da aplicação do método frente ao objetivo. O resultado final é obtido por meio de uma combinação linear ponderada dos critérios (*Weighted Linear Combination – WLC*) por meio dos pesos obtidos na etapa anterior, seguindo o esquema geral (Equação 3).

$$WLC = \sum_{i=1}^n (W_i * X_{ij}) \quad (\text{Eq. 3})$$

onde,

n: número total de critérios;

W_i: peso relativo da importância do i-ésimo critério;

X_i: é o valor do i-ésimo critério.

De acordo com Saaty (1996), como os valores dos julgamentos das comparações em pares são baseados em experiência ou em dados físicos, o AHP pode lidar tanto com aspectos qualitativos quanto quantitativos de um problema de decisão, o que seria um benefício da técnica. E possui uma larga aplicação entre a administração pública, indústria de manufatura, educação, indústria de óleo, eletrônica, aviação, medicina, telecomunicações, indústria de entretenimento, engenharias, entre outras (Russo e Camanho, 2015).

Em geotecnia, AHP tem sido usado no estudo da distribuição probabilística de parâmetros geotécnicos (Gong et al., 2006), na seleção do melhor método de mineração subterrânea (Jamshidi et al., 2009), na avaliação da estabilidade de taludes rochosos (Wang et al., 2013), na seleção do lugar mais indicado para construção de um aterro sanitário (Ghazifard et al., 2016), na identificação de parâmetros que afetam a liquefação de solos (Rajhans et al., 2013); no mapeamento da suscetibilidade a movimentos de massa (Kayastha et al., 2013), análise de desastres geológicos (Wang e Yi, 2009), avaliação do risco de inundação (Ghosh e Kar, 2018; Chakraborty e Mukhopadhyay, 2019), mapeamento da suscetibilidade a dolinas, em ambiente cárstico (Taheri et al., 2015; Calligaris et al., 2017), entre outros trabalhos.

2.3.2 Aprendizado de Máquina

A mineração de dados (*Data Mining* – DM) é definida como o processo de descoberta de padrões de uma grande quantidade de dados (Han et al., 2011) que podem ser aplicados no gerenciamento das informações, em tomadas de decisão, no controle dos processos estudados, entre outros. Ela é parte de um processo de pesquisa mais amplo denominado *Knowledge Discovery in Database* – KDD (Fayyad et al., 1996), que pode ser traduzido por Descoberta de Conhecimento em Base de Dados.

O processo de descoberta do conhecimento é constituído por fases que possuem uma ordem de execução (Figura 5), que se inicia com a obtenção e seleção dos dados, avança para a fase de pré-processamento e transformação dos dados, chega até a etapa de mineração dos dados e finaliza com a fase de avaliação do processo e assimilação do conhecimento (Fayyad et al., 1996)

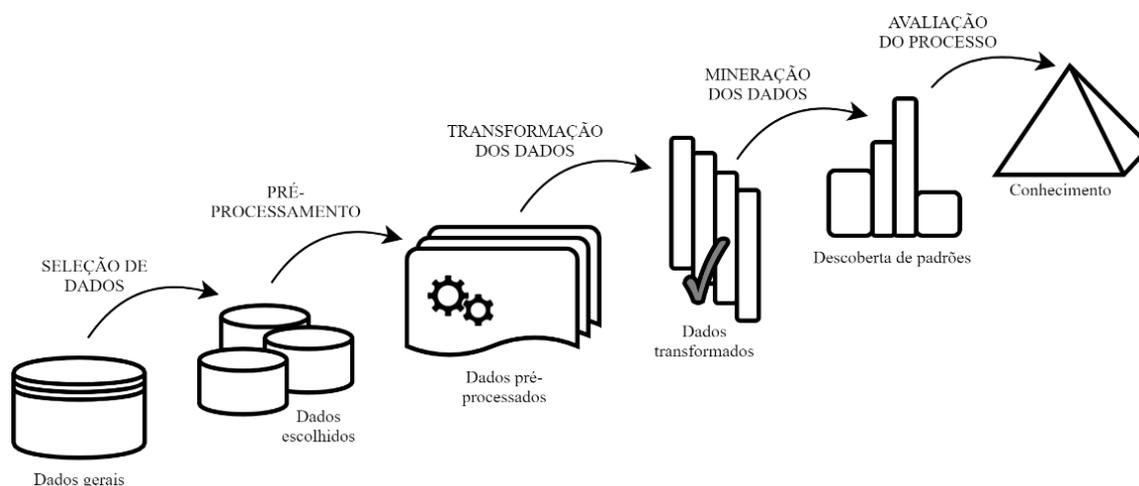


Figura 5: Fases do processo de busca de conhecimento utilizando mineração de dados.

As ações realizadas em cada fase do projeto são:

- Seleção dos dados:** elege os dados mais relevantes que podem ser efetivamente aplicados na técnica e que sejam úteis na descoberta de algum tipo de conhecimento;
- Pré-processamento:** limpeza de dados removendo dados inconsistentes, incompletos e fora dos padrões, preenchendo valores ausentes e/ou padronizando-os; organização de

um repositório único mantendo a consistência e coerência do banco, detectando redundâncias e valores conflitantes;

- c. **Transformação dos dados:** altera os dados em um formato apropriado para aplicação da técnica;
- d. **Mineração de dados:** consiste em identificar o tipo de técnica, aplicá-la, analisar e extrair resultados;
- e. **Avaliação dos padrões:** estudo dos resultados da aplicação da técnica para o objetivo proposto, validação e formatação dos resultados em gráficos, tabelas, mapas e relatórios;
- f. **Apresentação e assimilação do conhecimento:** expõe as descobertas obtidas, assimila o conhecimento adquirido para tomada de decisão, avalia as vantagens e desvantagens.

Silva et al. (2016) descrevem que o processo de DM busca descobrir padrões relevantes no banco de dados que sejam importantes para a compreensão das informações mais significativas na geração do conhecimento. Quando a DM acontece por meio de um método computacional que se baseia em algoritmos de previsão ou descrição, usando a experiência para melhorar seu próprio desempenho, conceitua-se esse processo como Aprendizado de Máquina (*Machine Learning* – ML) (Shalev-Shwartz e Ben-David, 2014; Mohri et al., 2018).

O aprendizado de máquina (ML) é um conceito advindo da Ciência da Computação que possui uma forte correlação com DM, entretanto, vários processos de DM não são resolvidos por ML e vice-versa. As principais diferenças, descritas por Mannila (1996), entre ML e DM são: (i) ML busca resolver problemas complexos, difíceis de serem solucionados por humanos, enquanto a busca por conhecimento usando DM possui objetivos mais modestos, em termos de complexidade, em que um analista de dados seria capaz resolver se tivesse tempo; (ii) para ML sempre existe algo a ser aprendido na análise de quaisquer dados que influem no problema, enquanto para DM os dados são as informações principais e não se supõe que exista qualquer estrutura oculta por trás deles; (iii) embora existam exceções, grande parte dos trabalhos que aplicam ML concentram o processo na etapa de aprendizagem, enquanto que a busca por conhecimento por KDD, onde se encontra a MD, é um processo constituído por várias etapas.

Desta forma, vale reforçar o conceito de Aprendizado de Máquina, que pode ser entendido como métodos computacionais, com regras e procedimentos, que usam as informações anteriores ou disponíveis, como entrada, para definir um padrão de comportamento, ou fazer previsões, e melhorar seu desempenho (Mitchell, 1997; Shalev-Shwartz e Ben-David, 2014; Mohri et al., 2018).

Os algoritmos constroem os modelos de ML a partir de dados amostrais de entrada (*inputs*, atributos ou variáveis) a fim de fazer previsões guiadas ou não, ou seja, com ou sem saída definida (*targets* ou rótulos). Como dados de entrada temos dois conjuntos: os dados estruturados e não estruturados (Silva et al., 2016). Os dados estruturados são aqueles que podem ser armazenados em estruturas tabulares que se relacionam. Dados não-estruturados são aqueles presentes em uma imagem, texto, vídeo ou sons – por exemplo, e não possuem estrutura de organização.

Existem duas formas de busca de conhecimento por ML. A primeira é a supervisionada: nesta busca já existe o conhecimento do que se está procurando (Mohri et al., 2018), ou seja, o conhecimento é advindo de um conjunto de exemplos (atributos) com saída definida (rótulos) que se comportam como dados de treinamento de um algoritmo e que, após o processo de aprendizagem, adquirem a capacidade de fazer previsões de rótulos para um novo conjunto de atributos. A segunda é não-supervisionada: nesta aplicação a busca por uma resposta é livre e espera-se que seja descoberta alguma informação significativa, ou seja, o conhecimento é advindo de um conjunto de atributos não rotulados, que após o processo de aprendizagem, podem apresentar relações e padrões entre os dados por meio de suas similaridades ou correlações entre os comportamentos (Silva et al., 2016).

O ML admite um conjunto de tarefas amplamente estudadas, alguns exemplos de tarefas são:

- a. **Classificação (*Classification*):** trata-se do processo de atribuir uma categoria a cada item ou evento. Nesta tarefa, o rótulo se apresenta como uma identificação da classe à qual o evento se relaciona, dentro de um número finito de classes existentes no domínio da análise;
- b. **Regressão (*Regression*):** trata-se de um processo que prevê um valor real de cada item ou evento. Nesta tarefa, o rótulo se apresenta como um número ao qual o evento se

relaciona, podendo ser contínuo ou discreto, dentro de um conjunto contínuos de valores para associação;

- c. **Agrupamento (*Clustering*):** trata-se de um processo que busca particionar um conjunto de itens ou eventos em subconjuntos homogêneos. Nesta tarefa, os atributos são um conjunto de dados e a saída é uma partição do conjunto de dados de entrada, definida por uma função de distância ou de similaridade. Geralmente usado para analisar um conjunto de dados muito grandes;
- d. **Redução da dimensionalidade (*Dimensionality reduction*):** trata-se de um processo que busca transformar uma representação inicial de eventos, banco de dados, em uma representação de menor dimensão, preservando algumas propriedades da representação inicial.

A tarefa que se enquadra na execução da proposta deste trabalho, visando alcançar os objetivos propostos, é a de Classificação. A Classificação é a tarefa mais empregada dentro do amplo espectro de aplicação de ML.

Dentre as técnicas ML existentes para aplicação em Classificação tem-se: *Árvore de Decisão (Decision Tree)*, *Floresta Aleatória (Random Forest)*, *Redes Neurais Artificiais (Artificial Neural Network)*, *Naïve Bayes*, *Support Vector Machine*, entre outros.

Cada uma das técnicas possui uma funcionalidade distinta, algumas são mais utilizadas para análise prévia do banco de dados, outras para o descobrimento de padrões, existem ainda aquelas destinadas a estimação ou a predição de valores baseado num padrão conhecido, e também para classificação.

As técnicas de Floresta Aleatória e Redes Neurais Artificiais, utilizadas nas experiências práticas deste trabalho, serão detalhadas a seguir.

A. Floresta Aleatória

As Florestas Aleatórias (*Random Forest – RF*) é um processo de aprendizagem baseado em um conjunto de árvores de decisão, cada qual construída a partir de um subconjunto aleatório de conjunto de treinamento. RF utiliza o método *ensemble* em que vários algoritmos de predição

mais simples são combinados para tornar o modelo mais robusto (Han et al., 2011). Assim, RF utiliza várias Árvores de Decisão a fim de obter um melhor desempenho preditivo. Uma Árvore de Decisão (*Decision Tree* – DT) consiste em uma coleção de nós internos (nó de decisão) e nós folhas (nó de respostas), sequenciados hierarquicamente (Silva et al, 2016) com o objetivo de criar modelos de aprendizado que representam regras de decisão. Assim, uma coleção de atributos, chamados de nós raiz, se ligam aos nós internos, por meio de ramos (valores de cada atributo) até que a condição de parada seja satisfeita, ou seja, chega-se ao nó folha.

A RF é, então, uma combinação de Árvores de Decisão, ou condições organizadas hierarquicamente, para produzir repetidas previsões do mesmo problema. Cada árvore depende dos valores de um vetor aleatório amostrado de forma independente e com a mesma distribuição para todas as árvores existentes na floresta (Breiman, 2001). A previsão da RF é obtida por maioria de votos sobre as previsões das árvores individuais, ou seja, será a classe selecionada pela maioria das árvores que constituem a floresta (Rodriguez-Galiano et al., 2014; Cutler et al., 2012; Biau, 2012). Em outras palavras, esse processo cria Árvores de Decisão a partir da seleção de uma amostra de dados de treinamento escolhida de forma aleatória, em seguida ele obtém a previsão de cada árvore e depois seleciona a melhor solução por votação (Figura 6).

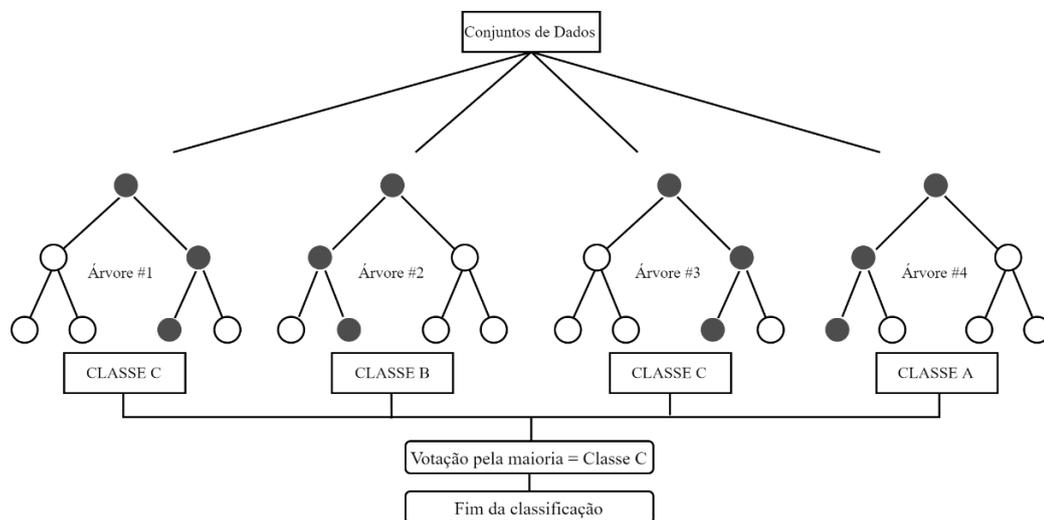


Figura 6: Exemplo do processo de Floresta Aleatória.

Breiman (2001) define RF como um classificador que utiliza um conjunto de árvores estruturadas $\{h(x, \theta_k), k = 1, 2, \dots, J\}$ onde J é o número de árvores na floresta e, os $\{\theta_k\}$ são vetores aleatórios distribuídos de forma independente e cada árvore lança um voto unitário para a classe mais popular na entrada x .

Para um melhor desempenho, há três parâmetros que podem ser ajustados:

- a. **Parâmetro “m”**: número de variáveis preditoras escolhidas aleatoriamente em cada nó. O valor de “m” é padronizado como sendo igual a \sqrt{M} , onde “M” é o número total de atributos na previsão para problemas de classificação. Para regressão, “m” é padronizado como sendo igual a $M/3$ (Breiman, 2001; Cutler et al., 2012);
- b. **Parâmetro "J"**: o número de árvores na floresta. Este número não deve ser muito pequeno, caso contrário, a estimativa *out-of-bag* pode ser instável e imprecisa (Cutler et al., 2012);
- c. **Tamanho da árvore (*tree size*)**: medido pelo menor tamanho de nó para divisão ou pelo número máximo de nós terminais. Taalab et al. (2018) explica que o aumento do número de árvores não melhora automaticamente o desempenho, mas pode aumentar a carga computacional. Quando as árvores são muito grandes, pode ser útil ajustar um limite de profundidade das árvores usando o parâmetro *tree size* para evitar o ajuste excessivo (Segal e Xiao, 2011; Cutler et al., 2012).

O algoritmo RF cria várias árvores a partir de uma seleção aleatória de dados de treinamento denominado *bootstrap*, e os dados que não são escolhidos são chamados de *out-of-bag*. Estes dados são extremamente úteis para erros de generalização e para determinar a importância de cada variável (Liaw e Wiener, 2002; Cutler et al., 2012).

O erro de generalização é estimado utilizando os dados *out-of-bag*. Durante o treinamento, o algoritmo pode utilizar os dados que não foram escolhidos para validar o modelo obtido pelos dados treinados, funcionando como uma validação interna (Cutler et al., 2012). Este procedimento é feito em cada árvore da floresta que não foi treinada pelos dados *out-of-bag*, em seguida é realizada uma previsão pela maioria dos votos de todas as árvores. A taxa do erro ou *error out-of-bag* (OOB) é dada pela comparação entre o valor encontrado nas previsões e o valor real das amostras *out-of-bag*.

O *error out-of-bag* é frequentemente usado para avaliar a precisão de uma floresta aleatória e selecionar os melhores valores dos parâmetros de ajuste (hiperparâmetros). Contudo, Janitzka e Hornung (2018) sugerem que o *error out-of-bag* pode se mostrar superestimado quando as

classes estão balanceadas, o banco de dados é muito pequeno, quando há uma correlação fraca entre as variáveis preditoras ou o número de variáveis preditoras é muito grande.

Para variáveis preditoras métricas e classificação binária, os autores sugerem utilizar a subamostragem estratificada como forma de contornar essa questão, por exemplo por meio da validação cruzada estratificada. A validação cruzada (*cross-validation* – CV) é amplamente utilizada quando a quantidade de dados para treinamento e testes for limitada, e o uso combinado da estratificação melhora ligeiramente os resultados (Witten et al., 2011). Entretanto, a CV despende mais memória e tempo de computação, especialmente com banco de dados de grandes dimensões (Janitza e Hornung, 2018).

Na CV é realizada uma divisão aleatória de um conjunto de dados em k subamostras (partições – *folds*) de tamanhos iguais, e assim é reservada uma parte dos dados para o treinamento e outra parte para os testes (Silva et al., 2017; Witten et al., 2011). O modelo é treinado usando $k-1$ *folds*, e a k -ésima *folds* é usada como um conjunto de teste. Este procedimento é repetido k vezes, e a cada vez uma partição diferente é usada como conjunto de validação. Como essa separação é feita de forma aleatória, não se pode ter certeza de que as amostras são realmente representativas do banco de dados. Por isso, é utilizado o procedimento chamado estratificação, que consiste em realizar uma amostragem de forma a garantir que cada classe seja devidamente representada tanto no conjunto de treinamento quanto no de teste (Witten et al., 2011).

Evidências teóricas e testes extensivos em vários conjuntos de dados diferentes, com diferentes técnicas de aprendizagem, mostraram que 10 é um número ideal de *folds* para obter a melhor estimativa de erro (Witten et al., 2011). Assim, um banco de dados é dividido aleatoriamente em 10 partes, nas quais a classe é representada aproximadamente nas mesmas proporções que no conjunto completo de dados (estratificação), uma parte é mantida e o esquema de aprendizagem é treinado nas nove partes restantes. A taxa de erro é calculada no conjunto de retenção, *fold* não utilizada no treinamento, e o procedimento de aprendizagem é executado um total de 10 vezes em diferentes conjuntos de treinamento.

Do ponto de vista computacional, a RF se destaca pelas seguintes características (Cutler et al., 2012): velocidade no treinamento e previsão; dependência de apenas um ou dois parâmetros de ajuste; possibilidade de obter a estimativa de erro de generalização; e possibilidade de medir a importância de cada variável para o desempenho do método. RF pode modelar interações

complexas a partir de variáveis contínuas ou categóricas, mesmo com valores ausentes, e ao mesmo tempo lida bem com variáveis preditoras irrelevantes (Cutler et al., 2012).

Como citado por Cutler et al. (2012), uma característica particularmente útil da RF é sua capacidade de classificar as variáveis preditoras em ordem de importância. Várias medidas, tais como *Gini Importance*, *Information Gain*, *Gain Ratio*, *Relief-F*, são usadas para estudar a importância de diferentes variáveis em RF (Quinlan, 1993; Kira e Rendell, 1992; Singh e Gupta, 2014).

O método mais utilizado em RF é o *Gini Importance* (Qi, 2012). O *Gini Importance* é derivado diretamente da função de divisão chamada *Gini Index*, que determinar qual atributo deve ser dividido durante a fase de aprendizagem da árvore (Menze et al., 2009; Qi, 2012). Em outras palavras, o *Gini Index* é a forma usada para decidir a divisão ideal (*split*) de um nó raiz e divisões subsequentes (Menze et al., 2009).

Para uma única árvore, o *Gini Importance* é definido como a soma da redução do *Gini Index* sobre todos os nós em que a característica específica é usada para dividir. Para uma floresta, o *Gini Importance* é definido como a média de sua importância entre todas as árvores da floresta (Cutler et al., 2012). Quanto mais o atributo diminui a impureza, mais importante é o atributo (Qi, 2012).

Breiman et al. (1984) descrevem a aplicação de árvores de decisão para tarefas de classificação e regressão desde o início da década de 80. Passando pela introdução do termo e do processo de RF por Breiman (1996, 2001) até os dias atuais, essas técnicas têm se mostrado como ferramentas robustas, precisas e bem-sucedidas para resolver inúmeras tarefas de aprendizado de máquina, incluindo, além de classificação e regressão, também estimativa de densidade e aprendizado semi-supervisionado (Criminisi e Shotton, 2013). E possuem um amplo uso em inúmeros campos, como mercado de ações (Patel et al., 2015), investigações bancárias (Arthi e Akoramurthy, 2018), bioinformática (Boulesteix et al., 2012), medicina (Masetic e Subasi, 2016), comércio eletrônico (Joshi et al., 2018), ecologia (na área agrícola e florestal) (Santoso et al., 2017), entre outros.

Em geotecnia, RF tem sido usado, por exemplo, na avaliação do potencial de liquefação de solos (Kohestani et al., 2015; Ahmad et al., 2020), no mapeamento do potencial de água

subterrânea (Naghbi et al., 2017), no mapeamento da suscetibilidade a movimento de massa (Tanyu et al., 2021), para estimar a resistência à compressão uniaxial e módulo de elasticidade de rochas (Matin et al., 2018) e, na predição dos parâmetros de resistência à compressão não confinada do solo (Ly e Thai Pham, 2020). Quando se trata de aplicações em espeleologia, os estudos se tornam escassos, podendo-se citar Zhu e Pierskalla (2016), Mohammady et al. (2019) e Elmahdy et al. (2020), que utilizaram RF para mapear locais com suscetibilidade a ocorrência de subsidência e sumidouros.

B. Redes Neurais Artificiais

A rede neural artificial (*Artificial Neural Network* - ANN) é um sistema computacional estruturado para receber informações, interpretá-las e tomar decisões apropriadas a partir de um treinamento. É um sistema dinâmico, não possui memória e a sua capacidade de aprender se faz com base na exposição de exemplos (Thomé, 2003). Silva et al. (2010) e Matssunaga (2012) apresentam algumas características da rede neural como: a adaptação por experiência; capacidade de aprendizado; habilidade de generalização; organização dos dados; tolerância a falhas (devido ao elevado número de interconexões), armazenamento distribuído (o processo é realizado de forma distribuída entre as diversas sinapses).

O estudo deste tema se iniciou a partir da observação das atividades cerebrais humanas, onde as entidades básicas – neurônios, interconectados em redes permitem a troca de informação entre eles (Haykin, 2004). O primeiro modelo matemático de um neurônio artificial foi desenvolvido por McCulloch e Pitts em 1943, que buscavam entender o funcionamento do cérebro (Kasabov, 1998). Para o treinamento de uma rede neural Donald Hebb propôs, em 1949, a primeira lei de aprendizagem, denominada regra de aprendizado de Hebb (Kriesel, 2005). A partir deste momento, várias redes neurais constituídas apenas por uma camada foram desenvolvidas, se destacando a Rede *Perceptron*, proposta por Frank Rosenblatt em 1958, que era a primeira rede capaz de adaptar os pesos internos dos neurônios para resolver uma classificação linear (Rauber, 1997).

Durante um tempo os trabalhos sobre redes neurais ficaram paralisado devido ao estudo de Minsky e Papert que no ano de 1969 apresentaram as limitações das redes neurais artificiais constituídas por somente uma camada (Kriesel, 2005). E, somente em 1986, com a publicação do método *backpropagation*, por Rumelhart, Hinton e Williams, que este tema voltou a aquecer

a comunidade científica. Com o método *backpropagation*, problemas não-linearmente separáveis poderiam ser resolvidos por meio da rede neural *Perceptron* multicamadas (Silva et al., 2010).

A ANN é um modelo útil para a resolução de problemas de forma semelhante às características principais do cérebro do ponto de vista do processamento de informação. A técnica computacional busca imitar o funcionamento do sistema nervoso do ser humano e suas unidades processadoras são denominadas neurônios artificiais, que se assemelham aos neurônios biológicos. Os neurônios artificiais buscam modelar tanto a estrutura quanto a funcionalidade de um neurônio biológico, e também a conectividade, a interatividade e, principalmente, a dinâmica operacional (Thomé, 2003).

A função de um neurônio artificial (Figura 7) é coletar os sinais de entrada ou *inputs* ($X_1, X_2...X_n$), ponderá-los por meio dos pesos sinápticos ($W_1, W_2...W_n$) e agregá-los de acordo com um combinador linear (Σ), levando-os a produzir uma resposta (saída do corpo celular artificial - u) após a introdução do limiar de ativação (θ). Neste momento, a função de ativação (g), recebe a resposta (u), e limita a saída ou *output* (y) a ter um valor dentro de um intervalo de valores a serem assumidos pela função.

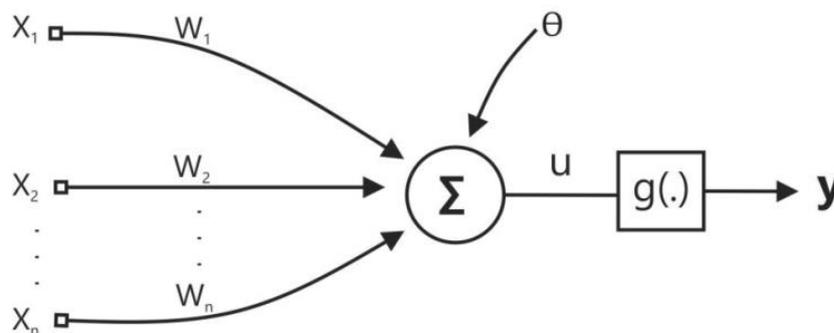


Figura 7: Neurônio artificial simples (Nola, 2015)

Silva et al. (2010) apresentam as descrições dos elementos que compõem um neurônio artificial. Os *inputs* são as variáveis assumidas em uma aplicação específica. Os pesos sinápticos permitem quantificar a relevância de cada variável. Eles são iniciados com valores randômicos entre 0 e 1, e durante o processo de aprendizagem são ajustados. O combinador linear tem a função de agrupar os produtos produzidos pela multiplicação dos *inputs* e pesos

sinápticos. O limiar de ativação tem a função de determinar o valor que deve ser atingido pelo combinador linear para que se estabeleça um valor de saída. A função de ativação normaliza a saída do neurônio para valores que possam ser interpretados. Elas podem ser dos tipos: degrau, degrau bipolar, rampa simétrica, logística, tangente hiperbólica, gaussiana e linear. E, por fim, os *outputs* são os resultados finais do processo que pode ser o produto final ou, a depender do tipo de rede, pode ser a entrada de outro neurônio da mesma rede.

A maneira pela qual os neurônios de uma rede neural estão estruturados está intimamente ligada ao algoritmo de aprendizagem a ser usado para treinar a rede (Haykin, 2004), e definem a arquitetura de uma rede.

Haykin, 2004 e Silva *et al.*, 2010 apresentam três tipos de arquitetura:

- a. **Feedforward de uma única camada:** nas redes unidirecionais os sinais são transmitidos em uma única direção, em direção à camada de saída. A rede é simples, estruturada por uma camada de entrada que se projeta para uma única camada de saída.
- b. **Feedforward de múltiplas camadas:** os sinais também são transmitidos em uma única direção, em direção à camada de saída, entretanto esta arquitetura possui mais de uma camada de nós computacionais, camadas ocultas. Os sinais de entrada são transmitidos para a segunda camada (primeira camada oculta) e os sinais de saída desta camada se configuram sinais de entrada para a próxima camada, e assim por diante até chegar à camada de saída.
- c. **Recorrente:** nas redes recorrentes os sinais podem ser transmitidos em ambas as direções através dos *loops* (camadas ocultas) na rede. Esta rede apresenta pelo menos uma realimentação durante o processo. Segundo Matssunaga (2012), a presença da realimentação tem um impacto significativo na capacidade de aprendizagem da rede e no seu desempenho.

Diversos são os modelos de rede propostos na literatura, cada qual advindo de uma linha de pesquisa diferente e visando um melhor desempenho na solução de um tipo específico de problema. Os tipos de redes neurais artificiais mais conhecidos são: *Perceptron*, que é o precursor do desenvolvimento das demais redes; e *Perceptron de Multicamadas (MLP, Multilayer Perceptron)* que é a mais amplamente utilizada. Todavia, existem outros tipos de

redes como: redes de função de base radial, *Adaline*, redes recorrentes de Hopfield, redes auto organizáveis de Kohonen, redes LVQ (*Learning Vector Quantization*) e *counter-propagation* e redes ART (*Adaptive Resonance Theory*).

A rede MLP tem como principais características: possuir no mínimo uma camada intermediária (oculta) de neurônios (Figura 8), ter uma função de ativação não-linear, e exibir alto grau de conectividade (Kovács, 2006; Haykin, 2004; Silva et al., 2010). Esse tipo de rede possui arquitetura *feedforward* e para treiná-la é utilizado o algoritmo *backpropagation*, de retropropagação do erro (Haykin, 2004).

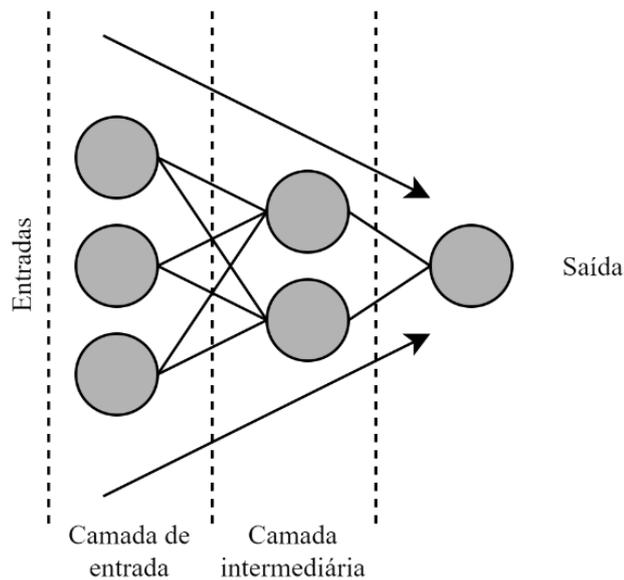


Figura 8: Representação de uma rede simples, tipo MLP, com uma camada intermediária.

O algoritmo *backpropagation* é um algoritmo de treinamento que se baseia na aprendizagem supervisionada, e é realizado por meio de duas fases (Silva et al., 2010, Haykin., 2004), são elas:

Fase 1: *Forward* ou propagação adiante. Nesta fase um conjunto de dados de treinamento é inserido na camada de entrada e o seu efeito se propaga, camada por camada, com um determinado conjunto de valores para os pesos sinápticos e os limiares, até chegar na camada de saída. O sentido de transmissão da sinapse é entrada→saída (Figura 9) Após obtenção dos dados de saída, estes são comparados aos valores

desejados (*targets*) e os erros produzidos pelos desvios entre resultado obtido e resultado esperado são calculados. Assim, a resposta esperada é subtraída da resposta obtida para produzir o sinal de erro, e os pesos sinápticos são ajustados de acordo com a regra da correção do erro.

Fase 2: *Backward* ou propagação reversa. Nesta fase o sinal de erro é propagado através da rede até chegar na camada de entrada, camada por camada, e os valores dos erros produzidos na etapa anterior são utilizados para ajustar os pesos sinápticos. O ajuste da camada de saída servirá como estimativa de erro para a camada anterior. Os ajustes das camadas intermediárias são realizados por meio dos erros de saída produzidos pelos neurônios da camada imediatamente posterior, que já foram ajustados. O sentido de transmissão da sinapse é saída→entrada (Figura 9).

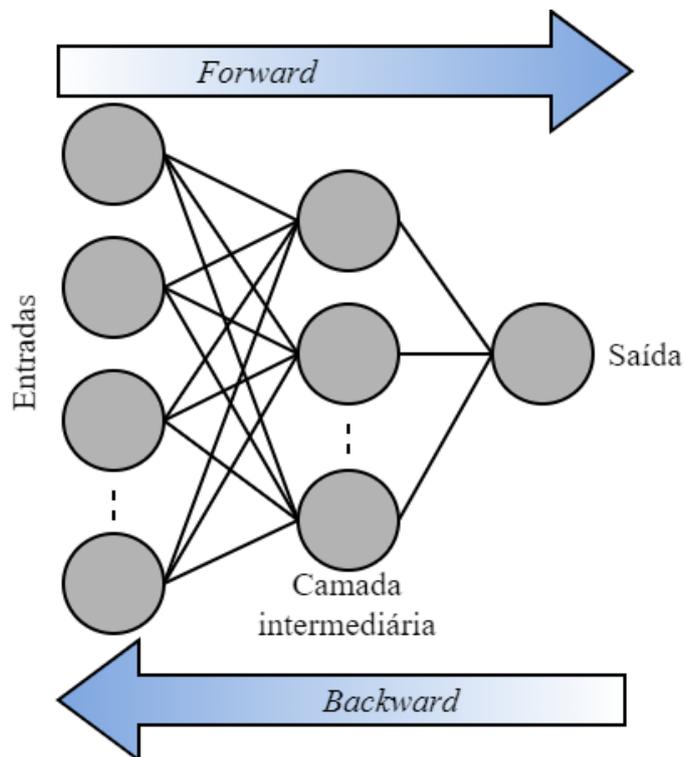


Figura 9: Sentido de transmissão do sinal no processo de aprendizagem usando o algoritmo *backpropagation*.

Silva et al. (2010) explicam que após a rede ter aprendido o relacionamento entre as entradas e as saídas, ela é capaz de generalizar soluções produzindo uma saída próxima daquela esperada a partir de quaisquer *inputs* inseridos em sua camada de entrada.

MLP é considerada uma das redes mais versáteis quanto à aplicabilidade devido a sua habilidade de aprender de experiências através do treinamento, e por esta razão ela alcança diferentes áreas do conhecimento (Haykin, 2004). Na década de 90, a utilização das ANN começou a ganhar forma e abriu-se um leque de aplicações em diversas áreas do conhecimento como: controle e automação, robótica, reconhecimento de padrões, previsão de eventos, energia, sistemas, fabricação, otimização, processamento de sinal, ciências médicas e ciências sociais/psicologia (Kalogirou, 2000).

Na área da Geotecnia, as redes neurais artificiais são utilizadas como forma de aplicação em diversos setores. Propriedades geotécnicas do solo, como a pressão de pré-adensamento (Çelik e Tan, 2005; Xia e Wu, 2014), a resistência ao cisalhamento (Kurup e Dudani, 2002; Kiran e Lal, 2016) e a permeabilidade (Sinha e Wang, 2008) podem ser estimadas usando ANN. Na área ambiental, esta mesma técnica foi utilizada no mapeamento de solos ácidos ricos em sulfatos que constituem um problema ambiental grave (Beucher et al., 2013). Sanchez (2009) aplicou ANN para modelagem de subpressões das fundações da barragem de Itaipu. Como informado anteriormente, desconhecem-se pesquisas que utilizem a técnica de Redes Neurais Artificiais na avaliação do potencial espeleológico de uma área.

CAPÍTULO 3

MATERIAIS E MÉTODOS

Para esse estudo foi elaborado, desde a fase de projeto, um roteiro de atividades que norteou os primeiros passos da pesquisa e amparou a fase de interpretação, aplicação das técnicas e obtenção de resultados. Para uma melhor visualização da organização desta pesquisa, apresenta-se o fluxograma ilustrado na Figura 10. Os procedimentos descritos nesse fluxograma podem ser agrupados em 3 conjuntos de atividades: os relacionados à fase inicial da pesquisa juntamente com a revisão bibliográfica, aqueles associados às atividades de análise dos dados e aplicação das técnicas e, por fim, a fase de elaboração dos artigos e finalização da tese.

3.1 PRIMEIRA ETAPA

Iniciou-se esta pesquisa com a definição da problemática que se pretendia estudar, de relevância científica, contribuição inédita e caráter inovador. Também estão incluídos a definição da hipótese, dos objetivos e a justificativa da pesquisa. Após a definição, aconteceu a escolha da área de estudo e foram levantados os trabalhos anteriores realizados na região relativos aos dados topográficos, geológicos, pedológicos e espeleológicos, inicialmente. Concomitantemente, foi escrita a revisão bibliográfica sobre o tema de forma a dar subsídio à etapa posterior, que é dependente principalmente da definição das variáveis que envolvem o problema e da técnica que modelam o problema.

Buscou-se escolher uma região dentro do Quadrilátero Ferrífero que tivesse a presença de litotipos ferruginosos, com pouco ou nenhum impacto antrópico, onde existisse grande quantidade de cavidades já prospectadas e que não estivesse inserida em processos minerários em fase de pesquisa ou licenciamento. Assim, a Serra do Gandarela foi escolhida como a região mais apropriada para o desenvolvimento deste trabalho. Demarcou-se um polígono de 7,2 km², totalmente inserido dentro do Parque Nacional Serra do Gandarela (*Serra do Gandarela National Park* - SGNP). O parque possui cerca de 230 cavidades em diversos litotipos, cadastradas no Cadastro Nacional de Informações Espeleológicas (CANIE), além de ser uma importante área de preservação ambiental dentro do IQ.

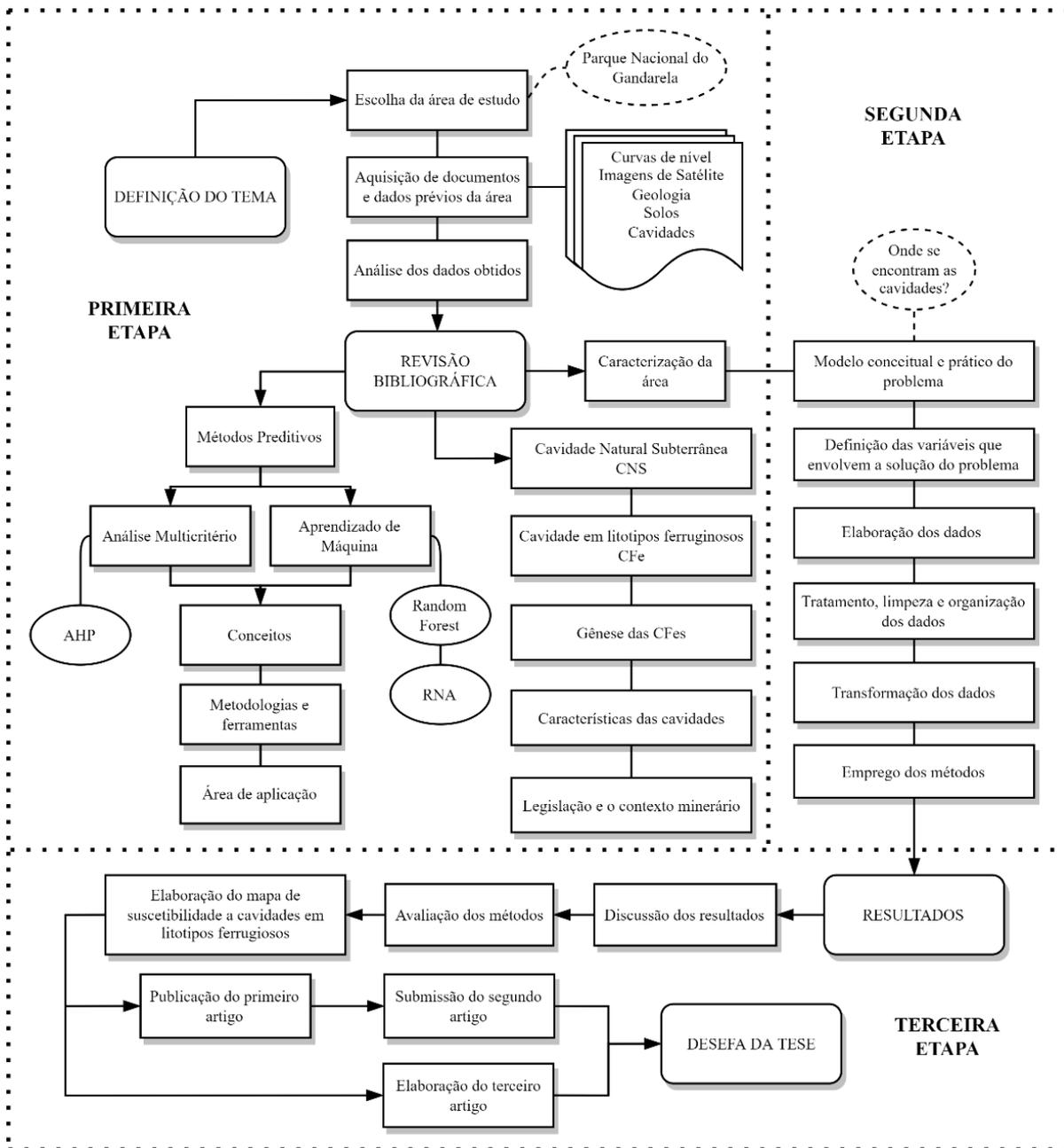


Figura 10: Diagrama representativo do processo de elaboração da pesquisa, por etapas.

Os dados prévios foram levantados para a caracterização da área de estudo e também para elaborar o banco de dados utilizado na aplicação das técnicas. Assim, os dados geológicos foram coletados a partir do Mapa Geológico de Ouro Preto, na escala de 1:50.000, de Lobato et al. (2005) e publicado pela CODEMIG. Como dado pedológico foi selecionado o Mapa de Solos – Bloco Acuruí, na escala de 1:50.000, de Shinzato e Carvalho Filho (2005) e pertencente ao Projeto APA Sul RMBH.

Os dados topográficos foram cedidos pela Vale e constaram de uma nuvem de pontos capturada por um sistema de varredura a LASER aerotransportado (aerolaser), mapeados em 2007, que foi utilizada para produzir um modelo digital de terreno de 10 m de resolução. As imagens de satélite foram adquiridas de dois diferentes satélites: da missão Sentinel 2A, coletada do banco de dados da Agência Espacial Européia (ESA); e da missão RapidEye, cedida pelo Instituto Estadual de Florestas (IEF). Também foram adquiridas pelo banco de dados do *Alaska Satellite Facility* (ASF) uma imagem radar capturada pelo satélite ALOS (*Advanced Land Observing Satellite*). Todos esses dados foram organizados, editados, tratados por meio dos softwares ArcGIS 10.1 (ESRI, 2012), e ENVI (ENVI, 2001).

Os dados de cavidades foram coletados, gratuitamente, na base de dados online do CANIE – gerenciado pelo CECAV (Centro Nacional de Pesquisa e Conservação de Cavernas). Esse banco de dados reúne informações de todos os tipos de cavidades distribuídos no território brasileiro. Para essa pesquisa foram selecionadas 55 cavidades que estavam inseridas dentro da área de estudo e tinham como litotipo: canga, canga/itabirito, formação ferrífera e itabirito.

O banco de dados foi composto por 384 instâncias com classes balanceadas, ou seja, 384 dados compostos pela junção de todas as variáveis estudadas, na posição de ocorrência ou não de cavidade. O posicionamento espacial de cada cavidade foi registrado pelas coordenadas geográficas da entrada da feição. O banco de dados foi formado a partir do pixel representativo da entrada da cavidade e cerca de 3 a 4 pixels adicionais. Esses pixels adicionais foram selecionados sempre acima da posição topográfica do pixel da entrada e limitando o mesmo. Ao final, 187 pixels representavam os lugares onde existem cavidades. Os pixels representativos dos locais onde não existem cavidades foram obtidos de forma aleatória em áreas em que não houve registros de cavidades e levando em consideração pontos coletados na expedição de campo realizada para reconhecimento da área de estudo. Desta forma, 187 pixels representavam os lugares onde não existem cavidades.

Para constituir o acervo bibliográfico desta pesquisa foram reunidos os livros, artigos, leis e trabalhos acadêmicos sobre os assuntos que envolvem os temas principais:

- Cavidades ferríferas: introdução ao tema, caracterização, gênese e evolução, legislação federal e estadual, e contexto minerário.

- Métodos para elaboração do mapa de suscetibilidade à ocorrência de cavidades ferríferas: Análise Multicritério por meio de AHP e Aprendizado de Máquina (ML) por meio de RF e ANN.

Houve também uma pesquisa específica sobre a área de estudo para sua caracterização e para o entendimento da ocorrência de cavidades na região.

3.2 SEGUNDA ETAPA

Esta etapa contemplou as fases do desenvolvimento da pesquisa para geração dos resultados preliminares a partir da definição das variáveis, buscando responder o seguinte questionamento: onde se encontram as cavidades ferríferas? Posteriormente, foram elaboradas todas as variáveis que poderiam ser utilizadas para solução do problema, o que constitui os resultados preliminares, e, em seguida, foi realizada a organização dos dados com a verificação da necessidade de tratamento e/ou limpeza. Na fase de transformação dos dados realizou-se a tabulação das variáveis em formatos apropriados para cada método aplicado.

As variáveis selecionadas para formarem o banco de dados de aplicação foram: razão de óxido de ferro (*Iron Oxide Ratio – IOR*), declividade (*Slope Gradient – SP*), curvatura horizontal (*Plan Curvature – PLC*), curvatura vertical (*Profile Curvature – PRC*), índice de vegetação por diferença normalizada (*Normalized Difference Vegetation Index – NDVI*), intensidade de lineamentos (*Lineaments Intensity – LI*), e o modelo HAND (*Height Above Nearest Drainage*). A seleção dessas variáveis está diretamente relacionada à gênese e às características físicas das cavidades ferríferas, além das características do meio físico onde elas são encontradas atualmente. A Tabela 7 apresenta um resumo das justificativas para a escolha de cada variável. Estas justificativas estão descritas de forma detalhada no Capítulo 4 (Artigo 1), juntamente com as referências para tais utilizações.

Tabela 7: Resumo das justificativas para a escolha de cada variável.

VARIÁVEL	JUSTIFICATIVA
Razão de óxido de ferro	Cavidades ferríferas estão associadas as cangas, formações ferríferas bandadas e itabiritos, especialmente nas porções mais ricas em ferro desses litotipos (Calux, 2013; Auler et al., 2019). O IOR busca mapear, remotamente, materiais geológicos superficiais ricos em ferro (Ducart et al., 2016; Gopinathan et al., 2020).

Tabela 7: Resumo das justificativas para a escolha de cada variável.

VARIÁVEL	JUSTIFICATIVA
Parâmetros morfométricos: Declividade, Curvatura horizontal e Curvatura vertical	A gênese das cavidades ferríferas está geralmente associada a processos erosivos de evolução das encostas (Piló e Auler, 2009). Os parâmetros topográficos são frequentemente utilizados nos estudos dos processos geodinâmicos das encostas (Pourghasemi et al., 2012). O processo erosivo, mecânico e geoquímico, é favorecido pelo aumento da declividade que altera a velocidade do escoamento superficial e subsuperficial (Summerfield, 2014). O estudo das formas de relevo é útil para avaliar a divergência ou convergência do fluxo de água superficial e subsuperficial ao longo do talude, mudanças na velocidade do fluxo e processos de transporte de sedimentos (Wilson e Gallant, 2000; Wysocki et al., 2011).
NDVI	Cavidades ferríferas podem estar associadas à quebra de relevo onde, normalmente, se observa uma zona de transição entre a vegetação escassa ou de campo rupestre ferruginoso, relativa à exposição das cangas, e a vegetação mais exuberante, desenvolvida sobre depósitos de tálus em rochas menos resistentes como os dolomitos (Piló et al., 2015b; Auler et al., 2019). O NDVI auxilia na identificação dos diferentes tipos de vegetação encontrados na região (Jensen, 2009).
Intensidade de lineamentos	A gênese das cavidades ferríferas está frequentemente associada a processos erosivos de evolução das encostas, que é facilitado pela infiltração das águas pluviais através das descontinuidades e a subsequente erosão química e mecânica do material menos resistente abaixo das cangas (Dias e Bacellar, 2021). Além disso, as cavidades ferríferas apresentam controle estrutural (Piló e Auler, 2009; Auler et al., 2022), observado nas paredes e tetos, e essas descontinuidades representam caminhos preferenciais para entrada de água no seu interior. A intensidade de lineamentos é o parâmetro mais indicado para representar a permeabilidade superficial do maciço rochoso (Singhal e Gupta, 2010; Florinsky, 2016) com baixa permeabilidade matricial, como nas cangas.
HAND	A gênese das cavidades ferríferas pode estar associada inicialmente ao ambiente hipogênico e saturado, que evolui para a fase vadosa (Auler et al., 2014; Calux et al., 2019). As cavidades ferríferas geralmente não apresentam fluxos de água permanentes em seu interior e, a interconexão hidrológica interna é rara (Auler et al., 2014). Devido à essas características, pode-se concluir que, atualmente, as cavidades ferríferas são encontradas acima da superfície freática regional (Auler et al., 2019). Entretanto, o uso do HAND visa mapear a profundidade da zona de saturação temporária no contato entre a canga e o itabirito (Auler et al., 2022), formando um interfluxo que alimenta os canais de drenagem intermitentes.

A construção de um adequado banco de dados composto por dados espaciais é definida pela uniformidade no sistema de projeção e na resolução dos rasters e, principalmente, na paridade das posições geográficas dos centros de cada pixel, pois é a partir da informação de cada pixel que é formada a matriz de dados de entrada para ambos os métodos estudados. Para isso, definiu-se o modelo digital de terreno (MDT) gerado pelo processamento da nuvem de pontos, pós tratamento, como modelo padrão. Este raster foi criado com resolução de 10 m (pixel) e sistema de coordenadas geográficas WGS 84, e todos os demais rasters foram padronizados com esta configuração.

Após a aquisição de todos os dados, foi necessário tratar os rasters que não estavam na mesma configuração do raster padrão. Esse tratamento constituiu de retirada dos pixels que estavam sobrando na borda do raster em decorrência do processamento em bases e softwares diferentes. Então, foi realizada a organização desses dados para formar o banco de dados. Para aplicação da técnica AHP, foram utilizados como dados de entrada os rasters, diretamente, sem nenhuma transformação. O processamento e a validação foram realizados no IDRISI Selva (Eastman, 2012) e no ArcGIS (ESRI, 2012).

A transformação dos dados para matrizes foi necessária para inclusão dos dados de entrada no *software* Orange 3.0 (Demsar et al., 2013). Este software foi desenvolvido pela Universidade de Ljubljana, Eslovênia, possui código aberto, e é específico para aprendizado de máquina com uma grande diversidade de ferramentas de análise em mineração de dados.

3.3 TERCEIRA ETAPA

A terceira etapa é composta pela elaboração dos artigos e a finalização da tese.

O emprego da técnica de AHP, a análise dos resultados desta aplicação, seguida da avaliação da utilização da técnica, culminaram na elaboração do mapa de suscetibilidade a ocorrência de cavidades ferríferas. Estes resultados subsidiaram a elaboração de um artigo científico (Capítulo 4), denominado *Multi-criteria Analysis for Mapping Susceptibility To Iron Formation Caves development in the Gandarela Mountain Range (MG), Southeast Brazil*, publicado em abril de 2021 no periódico *International Journal of Speleology*, Volume 50, Número 2, entre as páginas 173 e 187.

Os resultados obtidos por meio da aplicação de RF, juntamente com a avaliação da técnica e a comparação entre estes resultados e os obtidos no primeiro artigo, foram apresentados em um segundo artigo. Este artigo científico compõe o Capítulo 5, denominado *Mapping the susceptibility to the occurrence of iron formation caves using the random forest classifier*, foi submetido e aguarda publicação.

O terceiro artigo que compõe o Capítulo 6, envolve os resultados da aplicação de mais uma técnica de aprendizado de máquina, Redes Neurais Artificiais (ANN) além de um estudo comparativo entre técnica AHP e a técnica Floresta Aleatória, na avaliação da suscetibilidade a

ocorrência de cavidades ferríferas. Até o momento, este artigo não foi submetido a nenhum periódico.

CAPÍTULO 4

MULTI-CRITERIA ANALYSIS FOR MAPPING SUSCEPTIBILITY TO IRON FORMATION CAVES DEVELOPMENT IN THE GANDARELA MOUNTAIN RANGE (MG), SOUTHEAST BRAZIL

Os resultados e as discussões obtidos neste trabalho de doutorado estão aqui apresentados no formato de artigo científico, que foi aprovado e publicado no periódico *International Journal of Speleology*¹,

ABSTRACT:

In tropical regions, abundant in iron-rich geological materials, caves that are genetically and geographically associated with exploitable mineral deposits may develop. These caves have speleological relevance and are environmentally and legally protected in Brazil. Thus, for better planning of exploitation and environmental licensing, it is necessary to study the genesis and development of the iron formation caves seeking to preserve them without impeding the advancement of mining. This subject is complex, rarely studied, and few are the knowledges on alternatives to predict the occurrence of these caves. This gap justifies the development of research and products capable of assisting decision-makers, planners, and competent authorities in supporting the definition of target sites for speleological prospecting in the field. In this study, the prediction of the factors involved in the development of iron formation caves was evaluated producing a map of susceptibility in a GIS environment using fuzzy logic and an analytical hierarchic process (AHP). Therefore, the variables: iron oxide ratio, slope gradient, normalised difference vegetation index (NDVI), plan curvature, profile curvature, lineament intensity, and height above the nearest drainage (HAND) were selected. These variables were obtained by

¹ Artigo publicado no periódico *International Journal of Speleology*, 50(2), 173-187.

¹Autores: Iraydes Tálita de Sena Nola^a; Luis de Almeida Prado Bacellar^a

^a Núcleo de Geotecnia, Escola de Minas, Universidade Federal de Ouro Preto, MG - Brazil

processing of geospatial data from a region of the Gandarela Range (Minas Gerais, Brazil). The fuzzy logic and AHP techniques were applied, and for the validation of the results, a previously surveyed cave inventory was used. The results showed satisfactory performance of the map produced in predicting areas favourable to the occurrence of iron formation caves, presenting an area under the receiver operating characteristic (ROC) curve of approximately 0.85, which indicates a high prediction rate and validates the proposed method. Such results demonstrate that this susceptibility map was reliable and that the set of criteria and weights used were suitable for mapping areas favorable for speleological prospecting.

KEYWORDS: Iron Formation Caves, Susceptibility Map, Analytical Hierarchy Process, Fuzzy Logic, Brazil

4.1 INTRODUCTION

Among the environments favorable to the development of karst relief forms, those associated with the dissolution of limestone rocks are best known in a global context. However, also other types of rocks, such as quartzite, sandstone, granite, and shale, may develop karstic forms (Webb & Finlayson, 1984; Piccini & Mecchia, 2009; Sauro et al., 2014; Agapov et al., 2016; Holler, 2019). Karst features have also been described in iron-rich geological materials, usually in tropical regions, since the beginning of the 20th century (Gautier, 1906; Dixey, 1920; Jutson, 1914). According to Maurity (1995), karst features in iron formations have been found in India, Australia, Brazil and several countries of Africa.

In Brazil, the first records of this type of karst date from the 19th century (Souza & Carmo, 2015), but specific studies were only carried out in the middle of the 20th century (Guild, 1957; Simmons, 1963).

The Brazilian iron formation caves mainly develop in the Iron Quadrangle (IQ) and Carajás Mineral Province (CMP). In these areas, iron-rich geological materials occur, such as itabirites, banded iron formations, hematitic phyllites, and supergene materials such as ferricrete (“cangas”) (Auler & Farrant, 1996).

Similar to their counterparts in karst environments, the caves in iron formations have great speleological (Souza & Carmo, 2015), archaeological (Baeta & Piló, 2005; Carmo & Kamino,

2017) and palaeontological (Frank et al., 2012; Bittencourt et al., 2015) relevance. The iron formation caves also present peculiar biodiversity, containing extremely complex organisms, mainly troglomorphic, many of which have not yet been taxonomically described (Pipan & Culver, 2019); these comprise populations restricted to the humid and low luminosity underground environments (Ferreira, 2005; Souza-Silva et al., 2011).

There is a close genetic and geographical relationship between iron deposits and these caves (Souza & Carmo, 2015). Therefore, the occurrence of one or more iron formation caves in a mineral exploitation area may significantly affect mining planning, reducing economically minable iron reserves (Brandi et al., 2015). Thus, iron formation caves are often subject to legal restrictions in the face of mineral resource exploitation, which leads to intense economic and political pressure (Auler & Piló, 2015; Barbosa et al., 2019). Therefore, an important variable in mineral exploitation planning and environmental licensing is the identification and understanding of the genesis of iron formation caves, to meet the principles of sustainable development by preserving the caves without impeding the advancement of mining.

Some authors have investigated the genesis of these caves (e.g., Twidale, 1987; Grimes & Spate, 2008; Auler et al., 2014, 2019; Calux et al., 2019), but there are still many questions to be answered. Therefore, limited understanding of the genesis and, consequently, the geographic occurrence of these caves has not yet been addressed.

One of the techniques widely used in decision making for complex problems is multicriteria analysis through analytical hierarchy process (AHP). AHP combines information, knowledge, and expert judgment in an efficient and systematic manner. This combination helps meet the need for several types of assessments such as landslide susceptibility mapping (Kayastha et al., 2013), geological disasters (Wang & Yi, 2009), flood risk (Ghosh & Kar, 2018; Chakraborty & Mukhopadhyay, 2019), and in the case of karst relief forms, sinkhole susceptibility mapping (Taheri et al., 2015; Calligaris et al., 2017). However, studies applying this technique to map the susceptibility to iron formation caves development are still scarce in literature, which justifies further research on this subject.

Despite its wide application, one of the criticisms of AHP is the failure to address the uncertainties of pairwise comparison judgments (Pourghasemi et al., 2012). This failure is because decision-makers choose to express judgements in intervals rather than exact numerical

values (Najafi et al., 2014), which do not replicate human mental processes (Zhang et al., 2017). An alternative solution to this problem is the combination of the AHP technique and fuzzy logic (Yang et al., 2013).

Fuzzy logic enables the standardization of criteria by considering a continuous scale of data, making it more realistic than Boolean logic, which standardises the data by binary association (Ghosh et al., 2012). The use of fuzzy pertinence functions seeks to prevent classes from being grouped into fixed limits. This process is important for comparison of criteria by direct aggregation. The combination of AHP and fuzzy logic solves the problem of effective decisionmaking for multiple attributes (Yang et al., 2013), allowing more flexible combinations of weighted data.

Thus, the present work aims to map the susceptibility to iron formation caves development in Gandarela Range, a significant Brazilian mining region, through multicriteria analysis using AHP and fuzzy logic. This work is of great relevance as the numerous caves present in this mountain range have made it difficult or impossible to extract important deposits of iron ore, which represent one of the main commodities of the Brazilian economy.

4.2 STUDY AREA

The Gandarela Mountain Range (“Serra do Gandarela”) is located in the southern central portion of the state of Minas Gerais (southeast of Brazil), constituting the most environmentally preserved área of the Iron Quadrangle (Marent & Portilho, 2017). This region is part of the Serra do Gandarela National Park (SGNP), created to protect and preserve biological, geological, hydrological, and speleological diversity. The region also accommodates geomorphological complexes composed of mountains and plateaus associated with the ferricrete, including rupestrian grassland and remnants of the Atlantic Forest (Brazilian Government, 2014). The Iron Quadrangle is one of the largest mineral provinces in the world (Dorr, 1969), with vast reserves, especially of iron, gold, and aluminium.

A 7.2 km² polygon was selected as a study area in an NW oriented section of the mountain range where the largest amount of known natural iron formation caves in the region are concentrated (Figure 11). The vegetation includes semi-deciduous seasonal forests (representatives of the Atlantic Forest), gallery forest, and rupestrian grasslands (Oliveira et al.,

2005; Fernandes, 2016), which are associated with ferricrete exposure and characterised by herbaceous species typical of higher altitudes. The exposed geological units are part of metasedimentary sequence from the Paleoproterozoic, consisting of itabirites, dolomites, dolomitic itabirites, and hematitic phyllites (Dorr, 1969; Alkmim & Marshak, 1998). Caves typically develop in ferricrete, at the tops of mountains that configure as plateaus. Owing to the geological context associated with materials rich in iron and aluminium, mining companies have a great interest in the study region. Some interventions related to mineral prospecting, such as the exposed soil from trench opening and drilling, were present in the ferricrete plateaus before the creation of the SGNP.

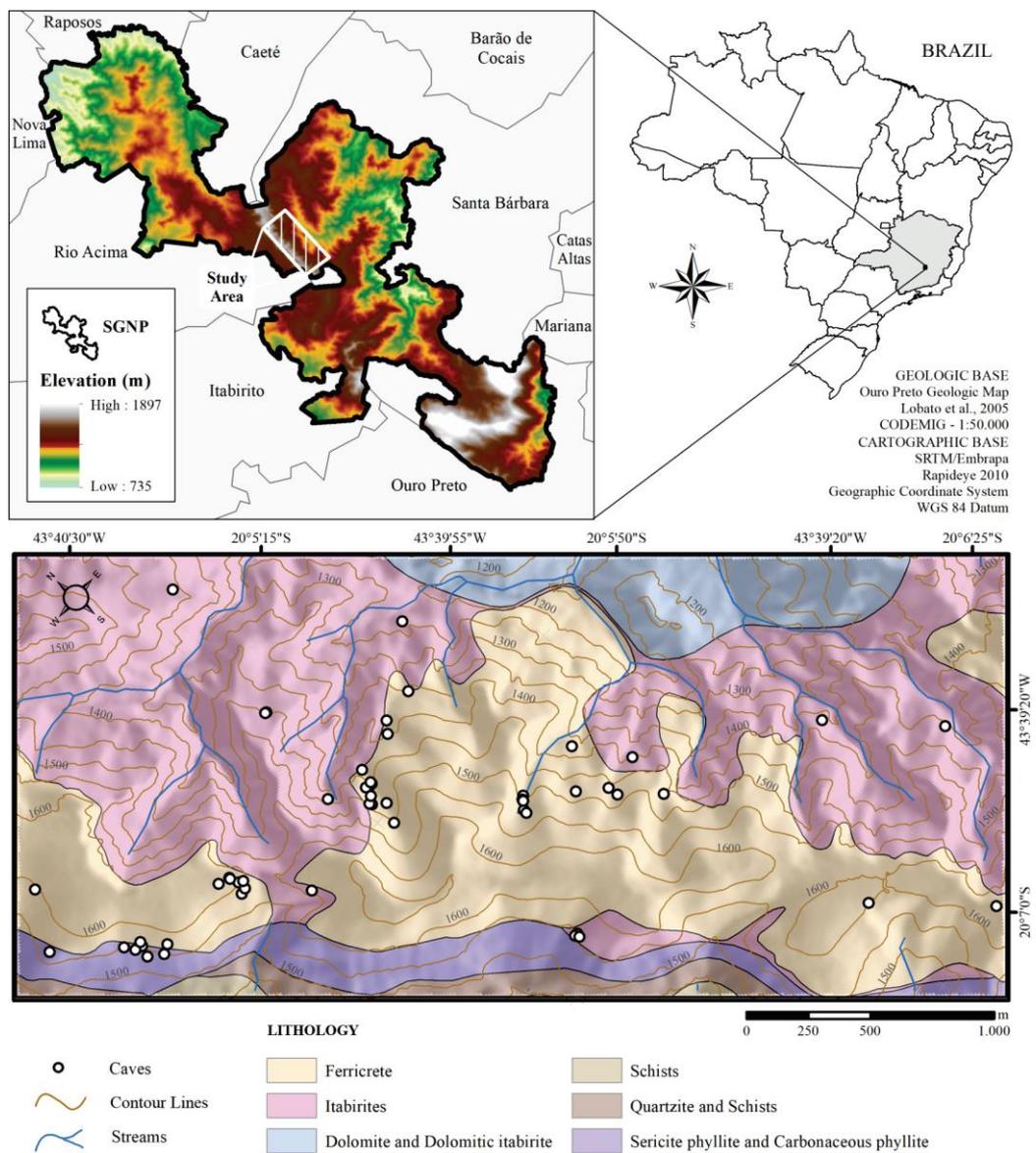


Figure 11: Location map of the study area indicating elevation and the geological units.

The natural factors that condition the genesis of caves in iron formations are still poorly understood. However, works developed at IQ and CMP have indicated certain indexes and parameters that influence the genesis of these caves.

Iron formation caves usually have small dimensions (Piló & Auler, 2009; Albuquerque et al., 2018), and are found at depths less than 5 m below the topographic surface (Auler et al., 2019). They develop concordantly with slope gradient (Auler et al., 2019; Piló & Auler, 2009), although at a gentler angle (Figure 12). The cave halls are small with irregular morphology (Fig. 10A, B), displaying pendants and pillars (Auler & Piló, 2005; Auler et al., 2014). The entrance of the caves is usually narrow and covered with vegetation (Simmons, 1963; Dutra, 2013). Fine and well-sorted ferriferous sediments and rock fragments from gravitational collapse processes are present on the floor (Fig. 10C) (Piló et al., 2015a; Dutra, 2017; Auler et al., 2019).

4.3 MATERIALS AND METHODS

4.3.1 Criteria influencing the development of iron formation caves

The iron formation caves in Brazil are associated with certain iron units, such as ferricrete, banded iron formations, and iron ores (Simmons, 1963; Dutra, 2013; Piló et al., 2015a; Auler et al., 2019), especially in the most iron-rich facies (Calux, 2013). Some caves develop at the interface between the ferricrete and banded iron formation, while others develop within the ferricrete formation (Auler et al., 2019). Therefore, an index that allows remote mapping of iron-rich geological materials will prove useful. There are several indices commonly used in digital image processing to identify iron-rich targets, as iron oxide causes absorption in the region of the spectral bands of blue and green and an increase in reflectance in the red band (Jensen, 2009). Among these, the iron oxide ratio (IOR) (Segal, 1982), representing the ratio between the spectral bands corresponding to red and blue, is often reported. This index has been used to discriminate iron oxides in lithological mapping (Ciampalini et al., 2013; Mazhari et al., 2017; Gopinathan et al., 2020), with good results, including under vegetation cover in the CMP region (Ducart et al., 2016).

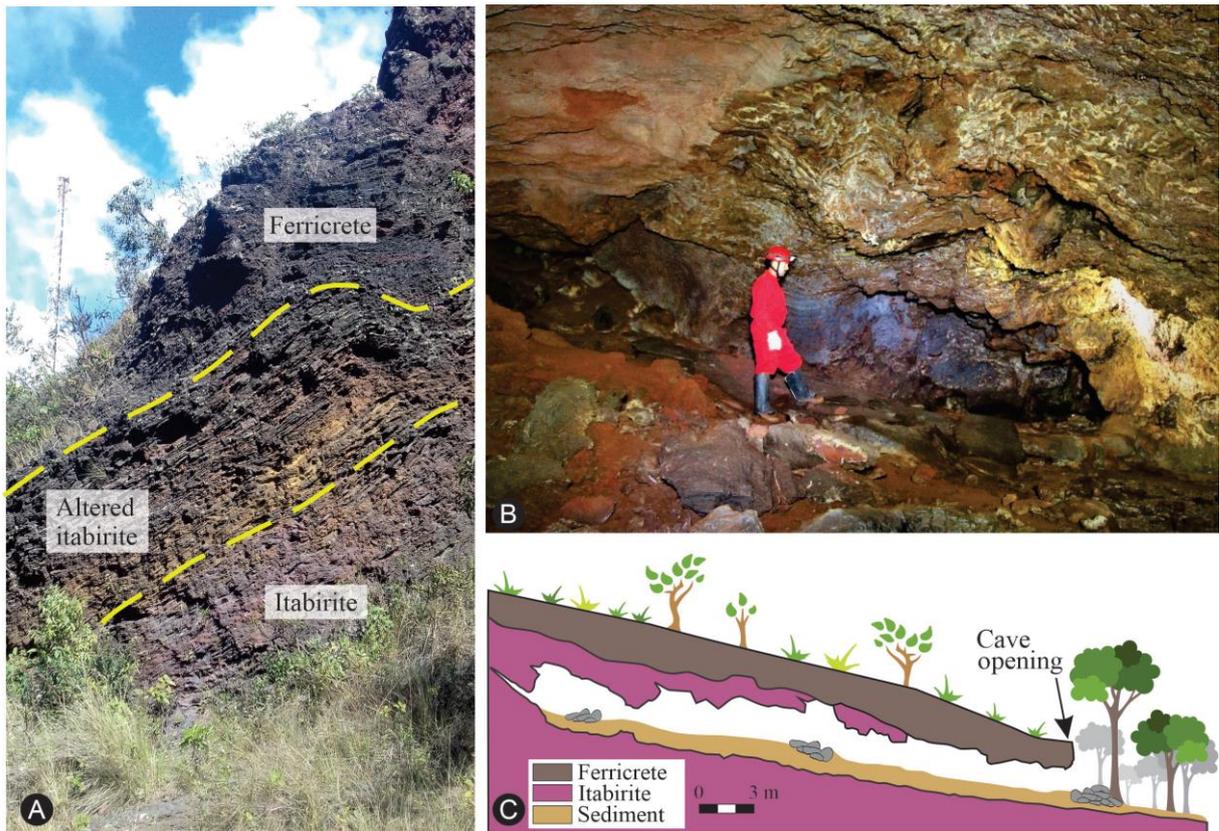


Figure 12: A) Weathering profile of an itabirite outcrop; B) Iron formation cave in FQ with typical irregular morphology; C) Schematic geological cross-section of a typical iron formation cave.

Morphometric parameters are of fundamental importance in several geodynamic processes, such as erosion (Bacellar et al., 2005), mass movements (Lee et al., 2004; Di Napoli et al., 2020), and even in the development of iron formation caves (Simmons, 1963; Piló & Auler, 2009; Auler et al., 2019). Among these parameters, the slope gradient (SP) and curvature are frequently used.

The increase in slope favours surface and subsurface water runoff and the consequent erosion process, both geochemical and mechanical (Huggett, 2011; Summerfield, 2014). Indeed, Leite (1985) shows that chemical leaching tends to increase with the intensity of interflow. Similarly, the curvature represents the shape of the terrain surface, allowing the assessment of the slope ability to control the flow of water through runoff (Wysocki et al., 2011) and subsurface runoff (Steenhuis et al., 2005). This runoff influences sediment distribution (Pahlavan Rad et al., 2016), vegetation (Florinsky & Kuryakova, 1996), and the intensity of geodynamic processes

(Pourghasemi et al., 2012; Arabameri & Pourghasemi, 2019). The curvature can be expressed both in terms of the plan and profile curvature (Ruhe, 1975). The plan curvature (PLC) depicts the shape of the terrain surface perpendicular to the direction of the maximum slope (Wilson & Gallant, 2000). It is associated with the ability of the water to converge or diverge on the slope (Wysocki et al., 2011). The profile curvature (PRC) depicts the shape of the surface parallel to the slope, i.e., in the direction of slope flow. It controls changes in flow velocity and sediment transport processes (Wilson & Gallant, 2000).

In this sense, Dutra (2013) states that the iron formation caves tend to develop on high to médium slopes, mainly in drainage headwaters with concave geometry in terms of plan and profile. Iron formation caves are frequent on the edges of plateaus and in small topographic relief breaks across stretches of slopes, as well as in the surroundings of drainage channels and lagoons (Simmons, 1963; Piló and Auler, 2009; Dutra, 2013).

Many authors (Calux, 2013; Piló et al., 2015b; Dutra, 2017; Calux et al., 2019) point out that the iron formation caves begin in a phreatic or transitional environment and evolve in the vadose zone. Despite being mostly humid, the caves do not present permanent water flows or significant water bodies (Dutra, 2013, Auler et al., 2014), and those with internal hydrological interconnection are rare (Calux & Cassimiro, 2015). Such findings indicate that currently, the caves are above the regional water table (Auler et al., 2019), which is often tens of metres deep (Calux & Cassimiro, 2015). However, after significant rainfall events, water inflow through the caves is observed, which percolates via fractures or bio pores (Dutra, 2013). In these cases, some flow may develop at the bottom of the caves as interflow (Freeze & Cherry, 1979), which could be added to the runoff and feed the temporary drainage channels identified downstream. Therefore, the temporary drainage channels tend to be related to the interflows identified in the caves. These interflows constitute temporary saturation zones above the regional phreatic surface.

The height above the nearest drainage (HAND) model proposed by Rennó et al. (2008) enables the assessment of the depth of temporary saturation zones. This model is created by extracting the vertical distance (relative difference) of each pixel from the digital elevation model (DEM) concerning the altitude position of the nearest drainage channel. This indicates the topographic normalisation of the landscape by the local drainage network, using the area of contribution of

this drainage as the process threshold. The values produced by the HAND model are directly related to the possible depth of the saturation surface feeding the drainage channels. Theoretically, the lower the HAND value, the greater the proximity of the terrain surface to the subsurface saturation level (Nobre et al., 2011), which in other situations represents the regional water table surface. This value justifies the application of the HAND model in studies related to soil hydrological characteristics (Gharari et al., 2011) in mapping areas subject to flooding (Nobre et al., 2016; Zhang et al., 2018), and morphodynamic processes (Kornejady et al., 2017).

Another relevant factor is the vegetation, which is directly associated with the bedrock and the topography (Lamounier et al., 2010). Thus, in iron-enriched regions where iron formation caves develop, like in ferricrete and itabirite outcrops, there is little to no soil formation and vegetation tends to be rarefied, classified as a ferruginous rupestrian grassland (Stávale, 2012; Piló et al., 2015a). In the context of FQ, rupestrian grassland is distributed in the highest parts of the relief.

Piló et al. (2015b) point out that canga substantially influences the floristic composition of rupestrian grasslands, including some endemic species (Mota et al., 2015). Iron formation caves are frequent in topographic breaks (Piló & Auler, 2009; Dutra, 2013; Auler et al., 2019), where talus deposits usually occur in association with large vegetation cover. Moreover, more developed vegetation patches inside plateaus with rupestrian grassland may represent collapsed caves or dolines (Pereira et al., 2012).

Therefore, indices that discriminate the various types of vegetation can be useful to distinguish areas that are susceptible to the development of iron formation caves. Among the most used indices, the normalised difference vegetation index (NDVI) stands out; this is a dimensional radiometric measurement given by the normalised ratio of multispectral images between the near-infrared (NIR) and red (Red) bands of reflectance (Jensen, 2009) (Equation 4). NDVI values indicate the density and vigour of the surface vegetation cover; the higher its value, the greater the amount of green vegetation (Holben, 1986). Valid NDVI values lie on a scale ranging from -1 to +1.

$$\text{NDVI} = (\text{NIR} - \text{Red}) / (\text{NIR} + \text{Red}) \quad (\text{Eq. 4})$$

Morphostructural lineaments are superficial linear features associated with some geological or geomorphological structures, such as faults, fracture zones, joints, ridges, and aligned valleys

(Clark & Wilson, 1994; Rogers & Engelder, 2004). They can be easily identified through aerial photographs or radar images (Sabins Jr, 1987; Casas et al., 2000). This variable is hydrogeologically important as it tends to indicate the path of water infiltration and percolation in low permeable rock mass. The factors considered most relevant for water circulation in rock mass with low primary porosity and permeability are intensity, aperture, and connectivity of discontinuities (Singhal & Gupta, 2010). As there are several patterns of discontinuities in the ferricrete region (Dutra, 2017), the connectivity can be considered good. Fracture aperture is a difficult parameter to measure in this region, as a single discontinuity may present a variation of this parameter in depth due to the continuous processes of dissolution and precipitation, common in the ferricrete (Ferreira, 2005; Spier et al., 2019). As there is a good correlation between lineament and fracture directions in the region (Dutra, 2017), the lineament intensity (LI) (National Research Council, 1996) becomes the only reliable variable for an indirect evaluation of superficial permeability (Florinsky, 2016). Therefore, this variable can be used in the study of groundwater recharge (Senanayake et al., 2016) and mapping groundwater zones (Srivastava & Bhattacharya, 2000; Magesh et al., 2012). In addition, tectonic planar structures, especially foliation planes and fractures, control the genesis and development of iron formation caves (Piló et al., 2015a). Structural conditioning is also marked by narrow passages between chambers, elongated chambers, conduit directions reflecting the main fracture orientations, and semi-parallel or parallel conduits (Piló & Auler, 2009; Dutra, 2013; Auler et al., 2019). The ferricretes in the FQ region are generally not permeable (Dorr, 1969; Souza, 2018), and discontinuities exert the main role for water circulation and spatial and temporal development of caves (Dutra, 2017). Therefore, any indicator that indirectly represents the intensity of fractures, such as lineament intensity, tends to have good significance for the genesis of caves.

4.3.2 Data processing

For the mapping of the susceptibility of iron formation caves, a model using fuzzy logic and AHP was proposed in this study, applied in a GIS environment, and later validated following the procedures illustrated in Figure 13. Among the criteria selected for the susceptibility analysis, the following seven were used: iron oxide ratio, slope gradient, NDVI, PLC, PRC, lineament intensity, and HAND.

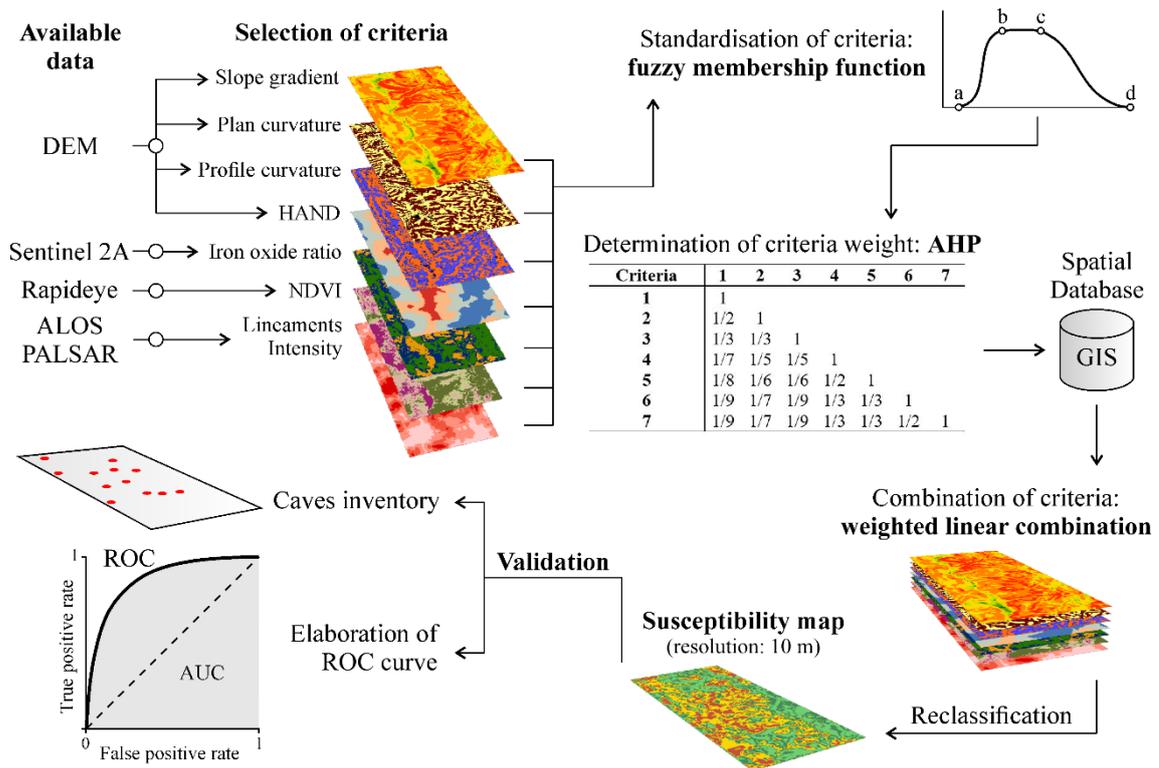


Figure 13: Flow chart for iron formation caves susceptibility assessment.

The geospatial data processing was performed with ArcGIS (ESRI, 2012), Terra View (INPE, 2019), ENVI (ENVI, 2001), and IDRISI Selva (Eastman, 2012) software. A 10-m cell was selected as the base resolution to maintain spatial consistency and facilitate operation between different rasters. Preparation procedures for each data layer are summarised below.

The iron oxide ratio (IOR) was determined based on satellite images from the Sentinel 2A mission, capturing high resolution multispectral bands. The scene used in this study corresponds to the processing level 1C, where the images are orthorectified with top atmosphere reflectance values (TOA) (ESA, 2015). The 10-m resolution bands 2 (490 nm) and 4 (665 nm) captured on 13/07/2016 by the Multi-Spectral Instrument (MSI) were used according to Equation 5. This scene was collected free of charge from the European Space Agency (ESA).

$$\text{Iron Oxide Ratio} = \text{band 4} / \text{band 2} \quad (\text{Eq. 5})$$

The slope gradient (SP), curvature, in terms of PLC and PRC, and the HAND were produced from a 10-m resolution DEM, derived from a cloud of points acquired by LIDAR. The treatment and filtering of the vegetation cover from the point cloud, as well as the creation of the DEM,

were processed using LAStools software, while the HAND model was produced using Terra View software (INPE, 2019).

The NDVI was determined using bands 3 (630 to 690 nm) and 5 (760 to 880 nm) captured by the satellites of the RapidEye mission, with the RapidEye Earth Imaging System (REIS) sensor. The orthorectified scene has a resolution of 5 m and was ceded by the Minas Gerais State Forest Institute (IEF). The resolution of the scenes was changed to 10 m in order to standardise the database.

In this study, the morphostructural lineaments were drawn based on a set of shaded relief rasters with different lighting directions, generated from the radar image (ALOS/PALSAR). This 12.5 m resolution image was captured by Advanced Land Observing Satellite (ALOS) with the Phased Array type L-band Synthetic Aperture Radar (PALSAR) sensor and obtained free of charge from the Alaska Satellite Facility (ASF) database. Therefore, the lineament intensity was obtained adding all the lineament lengths distributed over a defined area of radius = 150 m.

An inventory of iron formation caves was prepared to assist the process of evaluation and validation of results. Data were collected from the Brazilian Speleological Information Registry (CECAV, 2019) and 55 caves associated with iron formations were found in the study area.

The cave inventory was randomly divided into two groups: a training group (Group 1), with 70% of the iron formation caves used for the standardisation of criteria by fuzzy logic, and a validation group (Group 2), with 30% of iron formation caves, used for the validation of the result.

4.3.3 Standardisation of criteria using fuzzy sets theory

The inventory was standardised using fuzzy logic to attain the weighted linear combination (WLC) of the different criteria (Zadeh, 1965). This type of standardisation considers a continuous scale of the data, making it more realistic than Boolean logic, which standardises the data by binary association (Ghosh et al., 2012). The use of fuzzy pertinence functions seeks to prevent classes from being grouped into fixed limits, ensuring a continuous and more realistic variation (Schicker & Moon, 2012).

Fuzzy logic can be used in datasets of any measuring scale as the weighting of the criteria is entirely controlled by the expert, making it a simple technique to understand and easy to implement (Pourghasemi et al., 2012).

In this study, the original values of each criterion were replaced by continuous and dimensionless values, ranging from 0 (unfavourable) to 1 (favourable), indicating a symmetric sigmoid type fuzzy pertinence function (Brito et al., 2017), a function that best fitted the database.

The control points of the symmetric sigmoid function were determined from the mean and standard deviation of the values of each criterion. These values were obtained by overlaying the pixels of the iron formation caves (Group 1) on the respective pixels of the adopted criteria.

Table 8 presents the control points (a, b, c, d) obtained from the standardisation method used. The points a and d correspond to the mean minus standard deviation and mean plus standard deviation, respectively. Points b and c correspond to the mean values minus half the standard deviation and mean plus half the standard deviation, respectively. This standardisation technique aims to reduce human uncertainty in the decision making when considering the relevance of the values of each sub-criterion individually. Therefore, this step becomes essential in the preparation of data for AHP application in susceptibility mapping.

Table 8: Control points from the fuzzy relevance function to the criteria.

Criteria	Control points			
	$\mu - \sigma$	$\mu - \sigma/2$	$\mu + \sigma/2$	$\mu + \sigma$
	a	b	c	d
Iron Oxide Ratio (IOR)	0.553	0.617	0.746	0.811
Slope Gradient (SP)	29.673	38.809	57.081	66.218
NDVI	0.175	0.235	0.356	0.416
Plan Curvature (PLC)	-1.371	-0.807	0.321	0.885
Profile Curvature (PRC)	-0.740	-0.378	0.345	0.706
Lineament Intensity (LI)	7.825	8.956	11.218	12.349
HAND	103.727	139.853	212.103	248.228

4.3.4 Analytical Hierarchy Process (AHP)

AHP is a multi-criteria analysis method proposed by Saaty (1996, 2004), which uses the relative importance of criteria compared in pairs to give weight to each criterion in front of the problem exposed and thus, generates a solution for complex decisions.

The AHP can be divided into four stages (Tzeng & Huang, 2011): (1) creation of a hierarchical system from the decomposition of the problem into interrelated elements; (2) creation of the matrix of comparison in pairs between the criteria based on the attribution of the values of the scale of importance suggested by Saaty (1996, 2004), as presented in Table 9; (3) summarisation of the subjective judgments of the criteria and weighting of the relative weights from the calculation of the standardised autovector; and (4) aggregation of the relative weights to formulate the best alternative.

Table 9: Scale of importance between the criteria (Saaty 1996, 2004).

Intensity of importance	Definition	Explanation
1	Equal importance	Two activities contribute equally to the objective
3	Moderate importance	Experience and judgement slightly favour one activity over another
5	Strong importance	Experience and judgement strongly favour one activity over another
7	Very strong importance	Activity is strongly favoured and its dominance demonstrated in practice
9	Extremely important	The evidence favouring one activity over another is the highest possible order of affirmation
2,4,6,8	Intermediate values between the two adjacent judgements	When compromise is needed

4.3.5 Iron Formation Caves Susceptibility

In this study, the susceptibility to development of iron formation caves (CS) was integrated by a weighted linear combination of fuzzified criteria. The product is the iron formation caves susceptibility map that is represented by a raster file whose pixel values correspond to the CS values. This integration was performed pixel by pixel resulting in a continuous value ranging from 0 (lower susceptibility) to 1 (higher susceptibility). The Natural Break (Jenks) classifier

of the GIS environment (Calligaris et al., 2013; Myronidis et al., 2016) was used to determine the four classification ranges: very high, high, moderate, and low.

4.3.6 Validation

Validation is the most important step in the classification process, as it enables the analysis of the predictive capacity of the method (Hasekioğullari & Ercanoglu, 2012; Wu et al., 2018; Roodposhti et al., 2014; Kumar & Anbalagan, 2016). In this work, two forms of validation were used: the first consisted of superimposing the inventory on the map and visually evaluating the distribution of caves over the classes. The second consisted of receiver operating characteristic (ROC) analysis, which is widely used in the evaluation of classification methods (Nicu, 2018; Das & Pal, 2019; Kadam et al., 2019). The ROC analysis consists of correlating the rate of true positives (correctly classified cave data) and the rate of false positives (incorrectly classified cave data). From this operation, the area under the receiver operating characteristic curve (AUC) can be determined, with an AUC result close to 1 regarded as excellent. On the other hand, when the AUC value approaches its threshold, i.e., close to 0.5, it is evident that the method could not separate the desired classes. The success rate was calculated based on the AUC of the training dataset (Group 1) and the prediction rate was based on the validation dataset (Group 2).

4.4 RESULTS

Figure 14 presents the data layers with the results of the seven parameters investigated in the study area.

The weighting factors, evaluated pairwise, were applied to the seven criteria according to the importance of each criterion in relation to the susceptibility to development of iron formation caves, thus generating the comparison matrix (Table 10). The calculation of the eigenvalues of the comparison matrix determines the weights of each criterion (Table 10). Table 10 demonstrates that the most important parameter in relation to the occurrence of iron formation caves is the iron oxide ratio, followed by the slope gradient and NDVI. The least important criteria are lineament intensity and HAND.

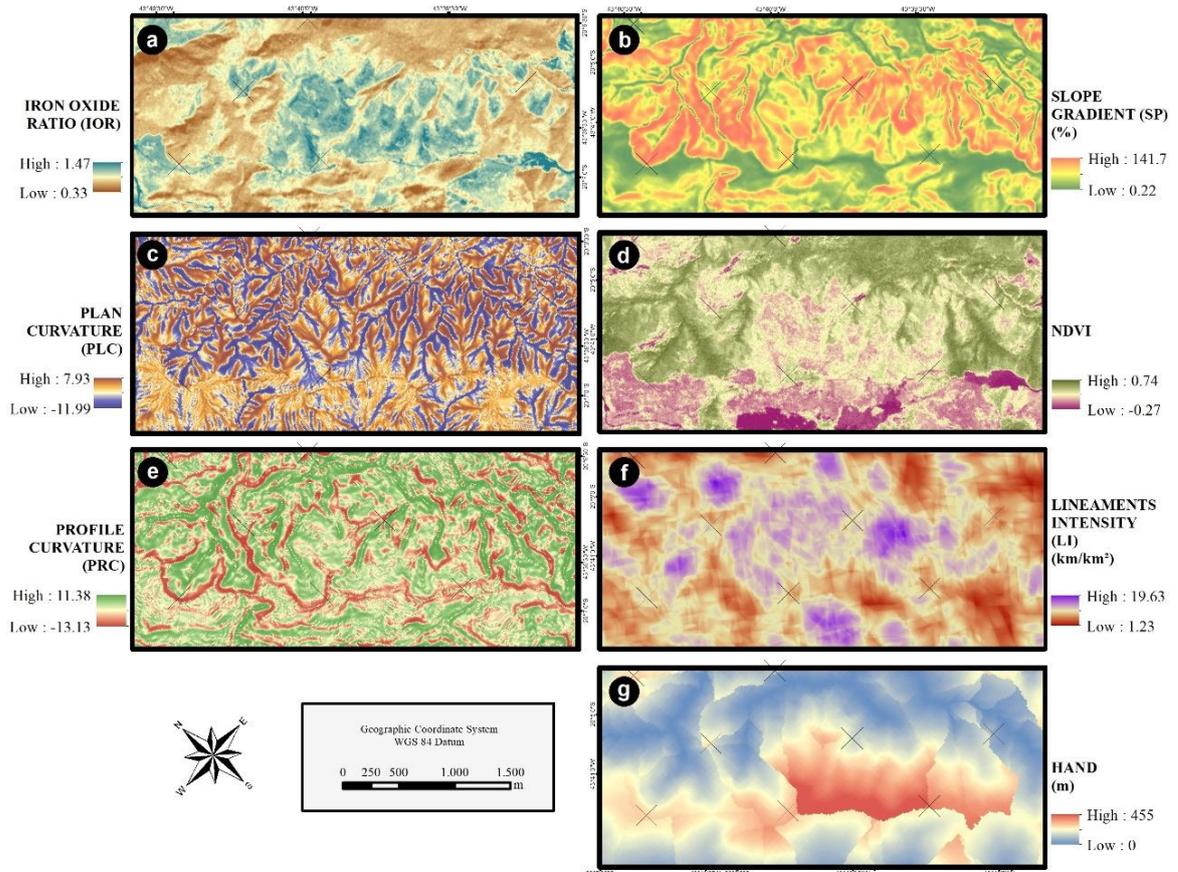


Figure 14: Data layers: (a) iron oxide ratio; (b) slope gradient; (c) plan curvature; (d) NDVI; (e) profile curvature; (f) lineaments intensity; and (g) HAND.

Table 10: Matrix for pair comparison, criteria weights, and consistency ratio.

Criteria	1	2	3	4	5	6	7	Weights
Iron Oxide Ratio (IOR)	1							0.367
Slope Gradient (SP)	1/2	1						0.257
NDVI	1/3	1/3	1					0.202
Plan Curvature (PLC)	1/7	1/5	1/5	1				0.065
Profile Curvature (PRC)	1/8	1/6	1/6	1/2	1			0.052
Lineament Intensity (LI)	1/9	1/7	1/9	1/3	1/3	1		0.031
HAND	1/9	1/7	1/9	1/3	1/3	1/2	1	0.025

The consistency ratio (CR) is used to evaluate the consistency of the classification by the weights in the AHP, for detecting contradictions in the determination of relative importance.

The CR value (Equation 6) is based on the consistency index (CI) (Equation 7) and the random consistency index (RI) (Table 11), developed by Saaty (1996, 2004).

$$CR = IC / RI \quad (\text{Eq. 6})$$

where,

$$CI: \lambda_{\max} - n / n-1 \quad (\text{Eq. 7})$$

λ_{\max} : main eigenvalue;

n: number of criteria of the problem;

RI: random consistency index for each n value.

Table 11: Random consistency index (RI).

N	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15
RI	0	0	0.58	0.90	1.12	1.24	1.32	1.41	1.45	1.49	1.51	1.53	1.56	1.57	1.59

The value of the limit for CR is 0.1 (Saaty, 2004). Larger values indicate that the weights should be revised. The CR value, in the study, was exactly 0.1, suggesting that the judgments applied in pairwise comparison are consistent.

The CS values were integrated by a linear weighted combination of the fuzzy criteria for each pixel. This value was given by the sum of the products of the weights assigned to each criterion, according to Equation 8:

$$CS = (IOR*0.367)+(SP*0.257)+(NDVI*0.202)+(PLC*0.065)+(PRC*0.052)+(LI*0.031)+(HAND*0.025) \quad (\text{Eq. 8})$$

The integration of the CS data resulted in a map of susceptibility to iron formation caves development, represented by four classes: very high, high, moderate, and low (Figure 15).

The results show that the very high class occupies 17.2% of the map and presents 35.1% of the mapped iron formation caves (Table 12). Similarly, the high class also presents 35.1% of the

caves, which occupies a larger area on the map (30.1%). The other classes, moderate and low, occupy 33.7% and 19.0 of the study area, respectively, covering 26.3% and 3.5% of the caves.

The very high class occupies isolated portions of continuous escarpments, with predominantly rupestrian grassland vegetation, and surrounded by the high class with shrubby vegetation. The moderate class occupies an area with high slopes associated with dense vegetation cover and concave plains. Flat or very steep terrain covered by dense vegetation, usually falls into the low class.

Table 12: Comparison between susceptibility classes and caves.

Classes	Breaks	Grid %	Caves %	FR ^(a)	NFR ^(b)
Low	0 – 0.17	19.0	3.5	0.18	0.09
Moderate	0.17 – 0.43	33.7	26.3	0.78	0.38
High	0.43 – 0.66	30.1	35.1	1.17	0.57
Very high	0.66 – 1.0	17.2	35.1	2.04	1

(a) FR – Frequency Ratio; (b) NFR – Normalised Frequency Ratio

The frequency ratio (FR) was calculated to evaluate the relationship between the classes and the occurrence of caves, which corresponds to the ratio between the relative frequency of cave occurrences and that of the percentage grid covered by each class. When this ratio is higher than 1, the relationship between caves and class is strong. When it is less than 1, the relationship is weak (Lee & Pradhan, 2007; Lee et al., 2012). Thus, as seen in Table 12, the FR of the very high and high classes present values of 2.04 and 1.17, respectively, indicating a strong relationship between the elements of the ratios. In contrast, the normalised frequency ratio (NFR) value of the high class is almost half of that for the very high class. Therefore, the definition of the class intervals proved to be adequate for mapping of susceptibility of iron formation caves development.

To validate the success rate and prediction of the map prepared in this study, two ROCs defined with a limit of 100 analyses were used (Figure 16). It is considered that the closer AUC value is to 1, the greater the model accuracy is (Kamp et al., 2008).

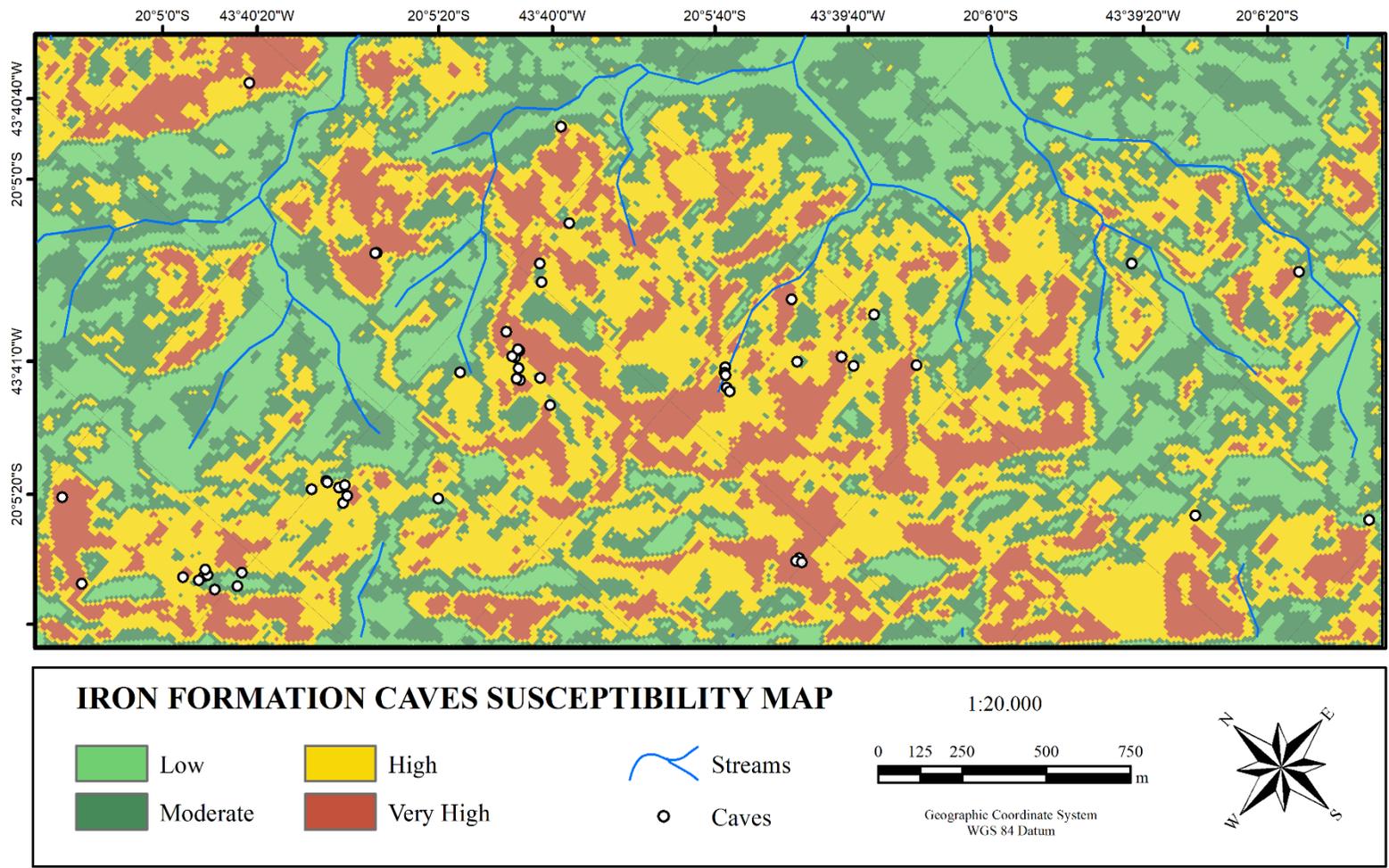


Figure 15: Susceptibility map of the study area produced by Analytical Hierarchy Process.

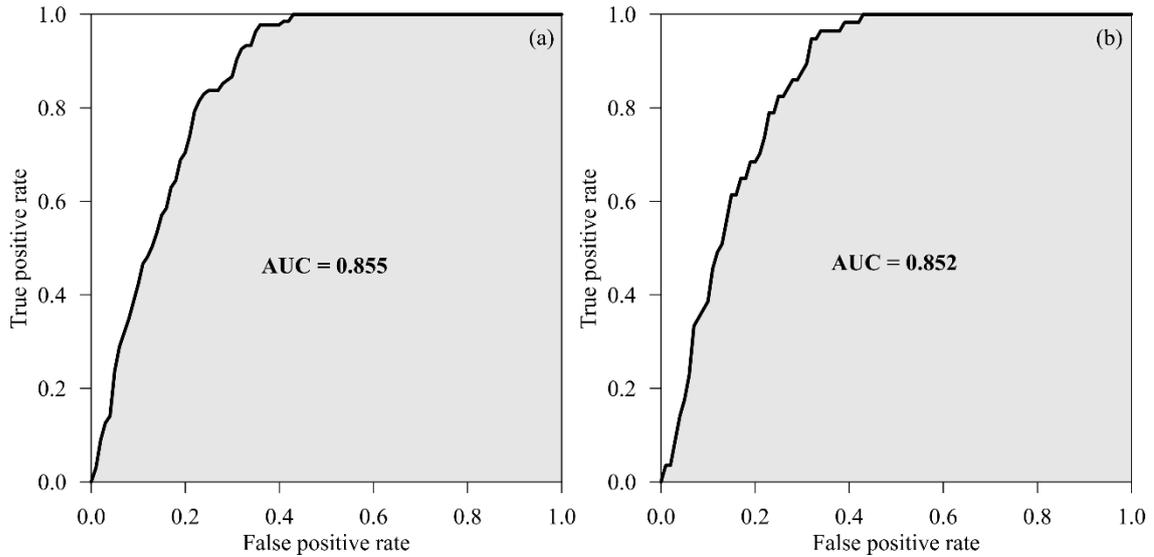


Figure 16: Receiver operating characteristic (ROC) curve analysis: (a) success rate curve using the training dataset; (b) predicted curve using the validation dataset.

It is observed that the AUC value for the success rate curve was 0.855 (Fig. 14a), while that for the predictive rate curve was 0.852 (Fig. 14b). The proximity between AUC values shows that the method exhibited a good and reliable performance (Rasyid et al., 2016). The two values of AUC (success rate curve and predictive rate curve) should be approximately equal; however, as the validation data are not used in modelling, the AUC value obtained using the validation data is usually lower than that for the success rate curve (Ngadisih et al., 2014). Our results demonstrate a good predictive ability of the proposed method to assess the susceptibility to iron formation caves development using AHP and fuzzy logic.

4.5 CONCLUSIONS

The occurrences of iron formation caves are closely associated with large iron ore deposits and this peculiarity makes the subject scarcely studied around the world. The detection of areas susceptible to the development of such caves is a very relevant theme and was the main objective of this work.

The method employed for assessing susceptibility (AHP) with standardisation of variables by fuzzy logic, has proved adequate. Among the seven indicators selected for the characterisation of susceptibility, the most relevant in descending order were: iron oxide ratio, slope gradient, and NDVI. The areas mapped as high and very high susceptibility occupy 47.3% of the land

and comprise 70.2% of the caves dataset. In comparison, the low susceptibility area occupies 19% of the land and 3.5% of the caves.

The validated results suggest a good performance of the method for rapid, inexpensive and remote-based prediction of areas favourable to the occurrence of iron formation caves. The proposed workflow can be considered as a tool capable of assisting decision-makers, planners, and competent authorities to support the identification of target speleological prospection sites. It is worth noting that the method has the potential to be used in two of the largest iron producing regions of the world (CMP and IQ), where such caves are frequent and interfere with the mining industry. This study is innovative in the use of this method, which presents as a limitation the lack of other data for an effective comparison. The aggregation of other data layers, such as aerogeophysical surveys, might further improve the quality of the results.

ACKNOWLEDGEMENTS

The authors would like to thank CAPES -Coordination for the Improvement of Higher Education Personnel, FAPEMIG – Minas Gerais Research Funding Foundation, CNPq - National Council for Scientific and Technological Development, and UFOP- Federal University of Ouro Preto for academic and scientific support. The Geotechnical Group (NUGEO) of the Escola de Minas - UFOP offered the laboratory facilities. The authors are also grateful to Vale SA for the opportunity to do this work and for providing data from the study region. The authors would like to thank the reviewers and IJS Editorial Board for the valuable comments, which have substantially increased the quality of the manuscript.

CAPÍTULO 5

MAPPING THE SUSCEPTIBILITY TO THE OCCURRENCE OF IRON FORMATION CAVES USING THE RANDOM FOREST CLASSIFIER

Os resultados e as discussões obtidos neste trabalho de doutorado estão aqui apresentados no formato de artigo científico, que foi submetido ao periódico *Computers & Geosciences*²,

ABSTRACT:

Caves and other karst features hosted by iron formations have often been described in various tropical regions of the globe. Many of these caves are protected by legislation thanks to their environmental, archaeological and paleontological relevance, even if the impact caused to the exploitation of the host iron ore-rich deposits are considerable. The subsurface identification of these caves their requires time-consuming and costly investigations. The aim of this study is to advance the analysis of automated methods to better understand the role played by natural factors in conditioning these caves, and to predict their spatial distribution. Thus, the influence of selected factors (iron oxide ratio, slope gradient, NDVI, plan curvature, profile curvature, lineaments intensity, and HAND) was assessed by means of the Random Forest (RF) algorithm. RF proved to be highly effective, with superior performance in identifying areas of greater susceptibility to the occurrence of caves, with success and prediction rates of 0.954 and 0.957, respectively. The NDVI (Gini = 0.137) and HAND (Gini = 0.123) factors were identified as those that most contributed to the prediction of the existence of caves, followed by slope gradient and iron oxide ratio. The findings of this research can be applied to the identification of favorable areas to speleological prospecting in iron formations, as it is a fast and economical method to remotely map the susceptibility to the occurrence of caves.

KEYWORDS: Iron Formation Caves, Susceptibility Map, Random Forest Classifier, Machine Learning, Brazil

² Artigo submetido ao periódico *Computers & Geosciences*.

² Autores: Iraydes Tálita de Sena Nola^a; Luis de Almeida Prado Bacellar^a

^a Núcleo de Geotecnia, Escola de Minas, Universidade Federal de Ouro Preto, MG - Brazil

5.1 INTRODUCTION

Iron ore is one of the most important commodities for industries in general, mainly for steel production (USGS, 2021). The high iron ore consumption by the steel industry has boosted exploitation in countries where iron ore reserves are world-class (Sonter et al., 2014), such as Australia and Brazil. In Brazil, there are two large iron ore-rich regions: the Carajás Mineral Province (CMP) in Northern Brazil and the Iron Quadrangle (IQ) in Southeastern Brazil, where banded iron formations or BIFs (jaspilites, itabirites) and other types of iron ore deposits are exploited (Rosière et al., 2006, 2008).

As in other tropical regions of the world, the morphological, geological and climatic conditions in these Brazilian regions favor the formation of caves in iron-rich geological units (Lefroy and Lake, 1972; Twidale, 1987; Vicat et al., 1995; Grimes and Spate, 2008). At present, ca. 1831 caves have developed in Brazilian BIFs (Auler et al., 2022), of which 63.0% are located in CMP and 22.1 % in IQ. Additionally, the Brazilian Speleological Information Registry (CECAV, 2021) shows that there are 57 caves developed in supergene deposits (bauxite and ferricrete). These iron formation caves (IFCs) are of archaeological and paleontological value, including a set of rare and even unique speleological features. They host specific forms of life, many of them exclusive to these environments, constituting a fragile and delicate ecosystem (Ferreira, 2005; Ribas and Carvalho, 2009; Piló et al., 2015a; Jaffé et al., 2016; Auler et al., 2019). IFCs are subject to negative impacts caused by the proximity to urban areas and mining companies (Piló et al., 2015b; Barbosa et al., 2019). Therefore, their protection and conservation are paramount and they are currently guaranteed by the Brazilian legislation. In face of a restrictive scenario imposed by the legislation in force, most of the mining projects have been affected by changes in mine planning and by the impossibility to extract iron ore from these localities (Auler and Piló, 2015).

Although IFCs have been object of study for decades (Simmons, 1963), the processes leading to their genesis have been characterized by several authors (e.g., Twidale, 1987; Grimes and Spate, 2008; Dutra, 2013; Auler et al., 2014; Auler et al., 2019, 2022; Calux et al., 2019), making it difficult to identify and map these subsurface features. Thus, studies that aim to understand the factors that condition development of these caves and to identify potential areas for their spatial distribution can make the optimization of speleological prospecting possible. Field prospecting in such areas is usually complex and expensive and even unsafe, and the type

of study proposed here can guide field survey to areas of greater susceptibility to the occurrence of IFCs.

Recently, Nola and Bacellar (2021) have satisfactorily mapped the susceptibility to the occurrence of IFCs in the Gandarela Range by applying Fuzzy Logic and the Analytical Hierarchy Process (AHP). Although AHP is widely used with satisfactory results in several areas (Kayastha et al., 2013; Minh et al., 2019; Chen and Dai, 2021), the judgment of the criteria used to formulate weights is subjective (made by a group of experts), adding a certain degree of uncertainty to the adopted weights applied to the conditioning factors and, consequently, to the results.

A predictive model with increasing application and that does not require experts' judgment to formulate weights to be applied to each variable is Machine Learning (ML) (Mohri et al., 2018). This model aims to explore by means of computational techniques a set of pre-existing information, in order to define a pattern of behavior, or make predictions, and improve its performance (Shalev-Shwartz and Ben-David, 2014; Mohri et al., 2018).

Among the ML techniques, the following stand out: Decision Tree (Akkas et al., 2015; Arabameri et al., 2021), Random Forest (Taalab et al., 2018; Norouzi et al., 2021), Artificial Neural Network (Kadavi and Lee, 2018; Xiong and Zuo, 2021), Naïve Bayes (Horrocks et al., 2015; Feng et al., 2018), and Support Vector Machine (Abedi et al., 2012; Cracknell and Reading, 2014). Although these techniques are commonly used in geosciences (Lary et al., 2015; Bergen et al., 2019), scientific studies that employ ML to predict the occurrence of IFCs are unknown.

Bergen et al. (2019) report that Random Forest is a ML algorithm that is easy to apply and interpret, as it requires the adjustment of a few parameters defined by the user. The aim of this study is to assess the performance of the Random Forest (RF) classifier in mapping the susceptibility to the occurrence of iron formation caves in the Gandarela Range, which is an important mining region inserted in the Iron Quadrangle (IQ) in Southeastern Brazil (State of Minas Gerais). It also aims to assess the importance of each variable in the RF performance. The results of this study will be compared those obtained by Nola and Bacellar (2021). Thus, it is expected that this study will help decision makers, planners, and competent authorities optimize speleological prospecting, making it more effective, faster and safer.

5.1.1 The study area

The IQ is located in the south-central portion of the State of Minas Gerais, in Southeastern Brazil, and constitutes the second largest Brazilian mineral province (Dorr, 1969), with large mineral reserves of iron, gold and aluminum. IQ is a geological and geomorphological complex controlled by major geological structures, such as anticlines and synclines, and by differential erosive processes (Varajão, 1991; Salgado et al., 2004). In the IQ mountainous regions there are rocks that are more resistant to erosion, such as the Moeda Formation quartzites and the Cauê Formation BIFs (Alkmim and Marshak, 1998).

The 7 km² study area is located in the southwestern portion of the Gandarela Range, where most of the IFCs already discovered in IQ are concentrated (Figure 17). It is located in the Serra do Gandarela National Park (Brazilian Government, 2014), one of the most preserved regions of IQ (Marent and Portilho, 2017).

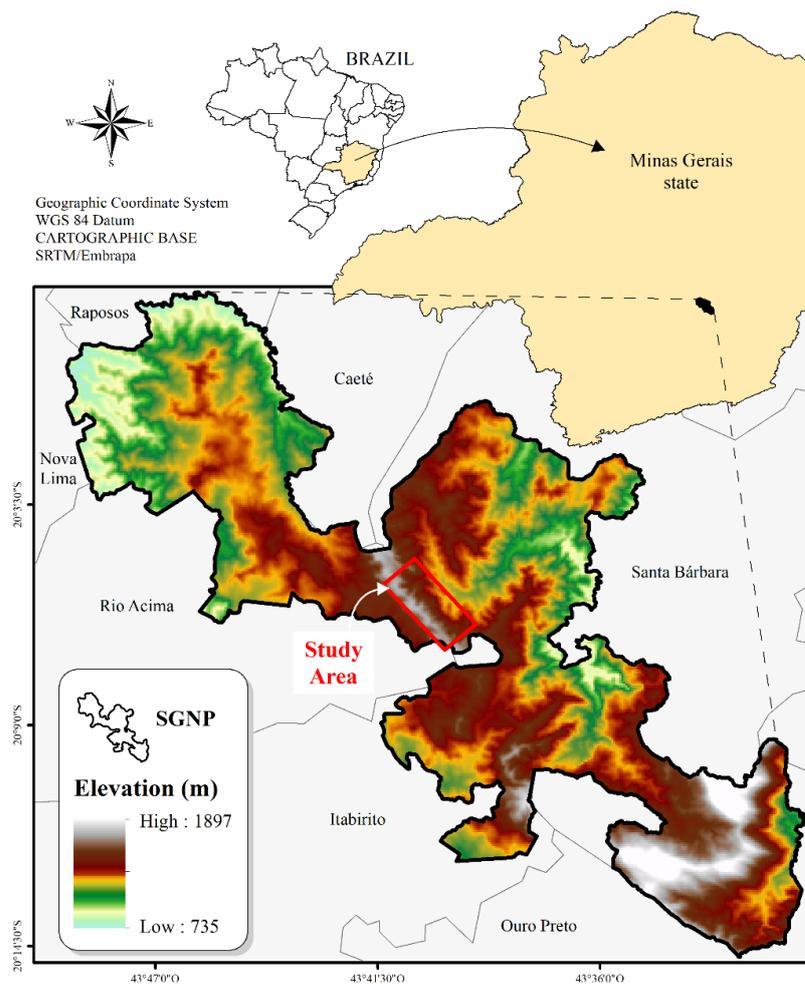


Figure 17: Location of the study area.

Itabirites, dolomites, dolomitic itabirites, quartzites, schists, sericitic and carbonate phyllites crop out in the study area, constituting a metasedimentary sequence dating from the Paleoproterozoic (Figure 18). The sequence is covered by supergene materials, especially ferricretes, locally known as “*canga*”, which develop especially on the itabirites (Dorr, 1969; Alkmim and Marshak, 1998). These ferricretes are of residual or colluvial nature and of varied thicknesses (Dias and Bacellar, 2021).

The residual ferricretes, also called structured ferricretes, result from the weathering of BIFs. Silica and mobile elements are leached, but the structures are preserved (Dorr, 1969, Souza and Carmo, 2015; Dias and Bacellar, 2021). Ferricretes form plateaus highly resistant to chemical and physical weathering (Salgado et al., 2004). Colluvial ferricretes, also called detrital ferricretes, result from the transport and deposition of fragments of itabirite and older ferricretes, bonded by ferruginous cements (Simmons, 1963; Dorr, 1969; Souza and Carmo, 2015). Detrital ferricretes generally lie on ramps located at lower altitudes (Souza and Carmo, 2015; Dias and Bacellar, 2021), disposed as layers that represent different episodes of colluvial-alluvial deposition (Dias and Bacellar, 2021).

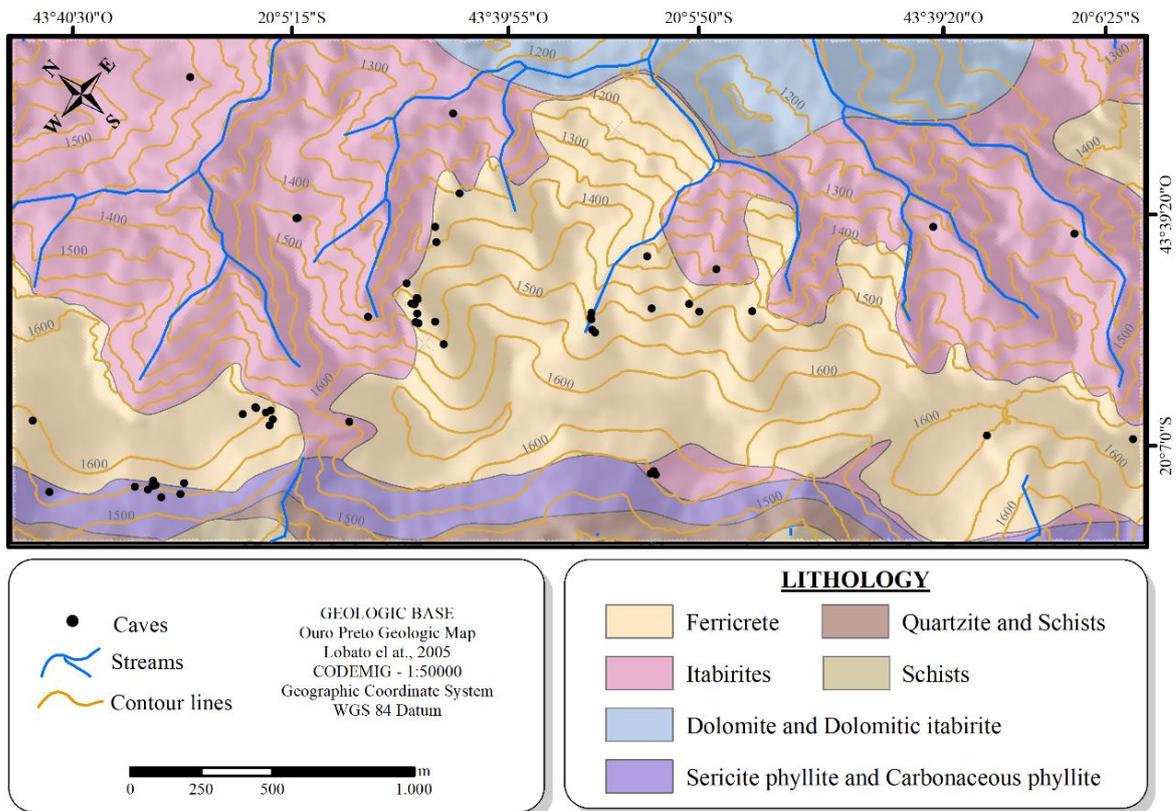


Figure 18: Lithological map and location of IFCs in the study area.

The vegetation is heavily influenced by the relief, which in turn is controlled by the rocky substrate (Salgado et al., 2004; Lamounier et al., 2010). Figure 19 shows a typical geological cross section of the study area, where the direct relationship between vegetation, relief, and rock substrate can be observed. The zoomed figure is the cross section of an IFC.

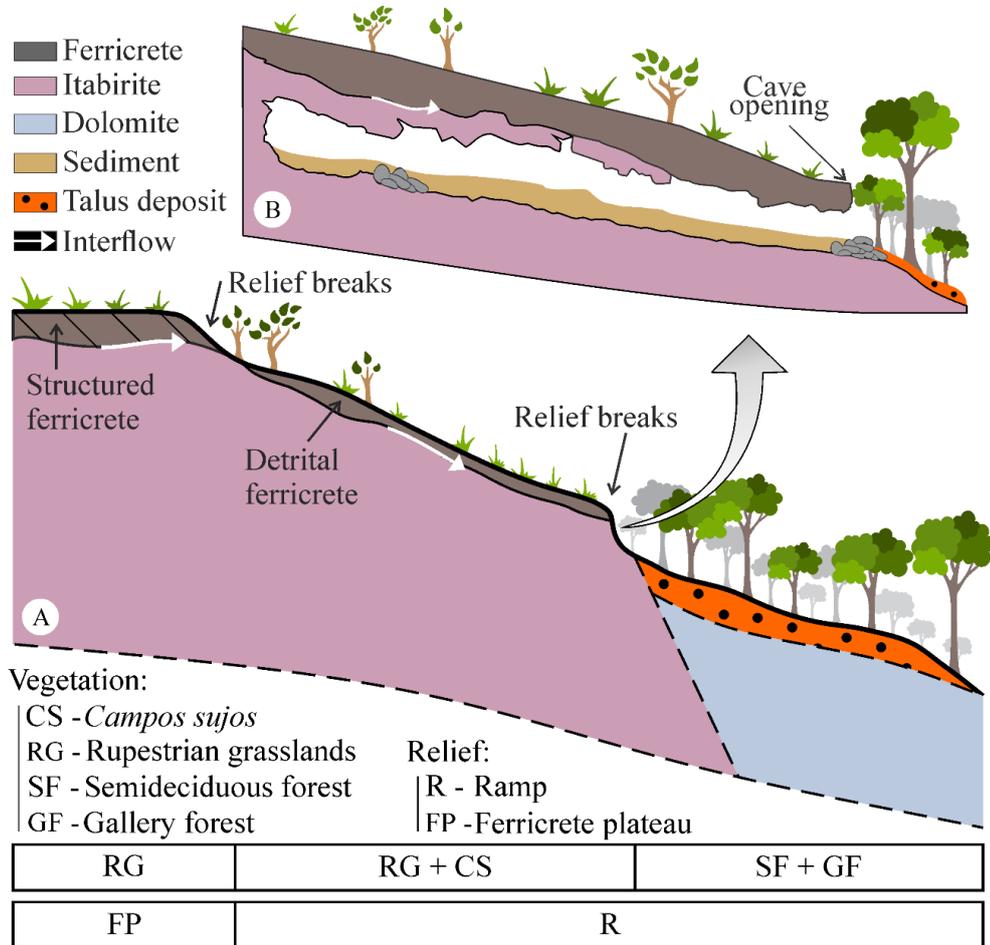


Figure 19: (A) Geological cross-section showing the typical relief and vegetation of the study area. (B) Zoomed figure showing a cross-section of an iron formation cave.

Plateaus occur in the higher portions (between altitudes 1660 and 1600 m); the slope gradients are gentle, less than 20% (Figure 19A). These plateaus are associated with the structured ferricretes and are characterized by rupestrian grassland covers (Fernandes, 2016; Marent and Portilho, 2017). In intermediate portions (between altitudes 1600 and 1450 m), continuous scarps form ramps with slopes gradients between 20 and 100%. The scarps are associated with detrital ferricretes and itabirites. Vegetation is an association of rupestrian grasslands and grasslands with the presence of bushes – referred to locally as *Campos Sujos* (Marent and Portilho, 2017). Relief breaks occur between the ramps and the hanging and deeply dissected

amphitheaters in itabirites and dolomites. They are marked by the abrupt change from rupestrian grassland to gallery forests or seasonal semideciduous forests of the Atlantic Forest type. The seasonal semideciduous forests occur in slopes gradients less than 20% and low altitudes (between 1450 and 1200 m). Dolomites predominate in these regions (Oliveira et al., 2005; Lamounier et al., 2010; Marent and Portilho, 2017).

A large part of the IFCs in Gandarela Range are positioned on the edges of the ferricrete plateaus or scarps and result from erosive retreat (Auler and Piló, 2015; Dutra, 2013; Auler et al., 2019; Dias and Bacellar, 2021) (Figure 19B). They develop in low, medium and high slope gradients, associated with relief breaks, and also around drainages and lakes (Simmons, 1963; Piló and Auler, 2009; Dias and Bacellar, 2021).

The caves can occur at the interface between the ferricrete and the weathered, less resistant BIFs or within the banded iron formation (Auler et al., 2019, 2022). Dias and Bacellar (2021) note that, although the contact between the ferricrete and the subjacent material (transition zone or altered rock) is an unsaturated zone, this contact constitutes water flow routes after heavy rainfall, indicating the existence of horizontal flows (interflows), as already observed by Ferreira et al. (2021) (Figure 19A, B).

IFCs are mostly small in size, with an average length of about 24.3 m and are located close to the surface, which indicates an extremely thin roof, less than 5 m in thickness (Auler et al., 2019). They have an ascending slope and profile (Figure 19B), following the dip of the slopes (Dutra, 2017), and the structural control is evidenced by discontinuities limiting their walls and ceilings (Piló et al., 2015b). The accumulation of ferricrete fragments (sediments) on the IFC floor is very common and results from internal ruptures conditioned by discontinuities (Piló et al., 2015a; Auler et al., 2019).

The entrance to the caves is usually narrow and covered with vegetation (Simmons 1963; Dutra, 2013) (Figure 19B). Halls are small and with irregular planimetry; the larger ones are characterized by the presence of irregular conduits connected to other smaller, younger ones (Piló et al., 2015a). They may have pillars, pendants, canaliculi and skylights. Speleothems are identified as coralloids with nodular, globular, botryoidal or coral-like morphology, consisting of iron and aluminum oxides and hydroxides, phosphates and sulfates (Silva and Santos Junior,

2015; Figueira et al., 2019). In general, speleothems in IFCs are smaller than those found in carbonate rocks.

5.1.2 Genetic Factors

Nola and Bacellar (2021) observed that IFCs develop by means of complex interactions between several factors. They also mention that the combination of some geological, morphological and hydrological particularities via Fuzzy Logic and the Analytical Hierarchy Process (AHP) can be useful to map the susceptibility to the occurrence of these features.

In this study, the same conditioning factors used by Nola and Bacellar (2021) were selected to build a processing database. These factors are seven: iron oxide ratio (IOR), slope gradient (SP), plan curvature (PLC), profile curvature (PRC), Normalized Difference Vegetation Index (NDVI), lineaments intensity (LI), and Height Above Nearest Drainage (HAND). Data obtained from geophysical surveys, such as ground-penetrating radar (GPR), resistivity and microgravimetry, can be useful to identify and delimit caves (Bishop et al., 1997; Conejo-Martín et al., 2015; Cardoso et al., 2018). However, such methods are more suitable for detailed speleological prospecting and are impracticable and very expensive for prospecting at a regional geological scale. Therefore, Nola and Bacellar (2021) proposed factors that were easily available. The weights adopted by these authors for each of the factors in the AHP analysis are shown in Figure 20.

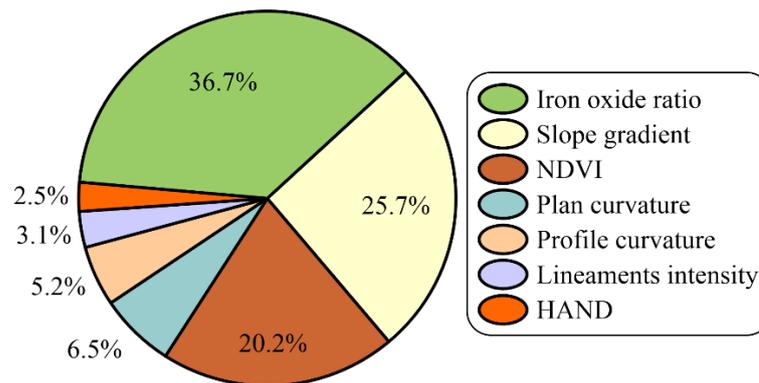


Figure 20: Weights adopted for each conditioning factor in the AHP analysis.

The selection of these factors is related to the genesis and physical characteristics of IFCs, in addition to the characteristics of the physical environment where they are currently found. Table 13 presents the minimum and maximum values of each conditioning factor adopted in this study

and the reasons for choosing each one. It was elaborated on the basis of the general information presented by Nola and Bacellar (2021).

Table 13: Main reasons for choosing the conditioning factors adopted in this study.

FACTORS	MIN – MAX*	REASONS
Iron Oxide Ratio (IOR)	0.33 – 1.47	IFCs are associated with ferricretes (<i>cangas</i>), banded iron formations and itabirites, especially in the iron-richer portions of these lithotypes (Calux, 2013; Auler et al., 2019). IOR allows the remote mapping of surface iron-rich geological materials (Ducart et al., 2016; Gopinathan et al., 2020).
Topographic parameters: Slope gradient (SP), Plan curvature (PLC) and Profile curvature (PRC)	<p>SP 0.22 – 141.7 %</p> <hr/> <p>PLC -11.99 – 7.93</p> <hr/> <p>PRC -13.13 – 11.38</p>	<p>The genesis of IFCs is generally associated with erosive processes that take place during slope evolution (Piló and Auler, 2009). Topographic parameters are frequently used in slope geodynamics studies (Pourghasemi et al., 2012). Erosion, both mechanical and geochemical, is favored by the increase in slope gradient that alters the velocity of surface and subsurface runoff (Summerfield, 2014). The study of slope shapes is useful in assessing the divergence or convergence of surface and subsurface water flow along the slope, changes in flow velocity, and sediment transport processes (Wilson and Gallant, 2000; Wysocki et al., 2011).</p>
Normalized Difference Vegetation Index (NDVI)	-0.27 – 0.74	IFCs are usually associated with relief breaks, where there is usually a transition zone between sparse vegetation and rupestrian grassland, related to the exposure of ferricretes, and more exuberant vegetation, developed on talus deposits on less resistant rocks, like dolomites (Piló et al. 2015b; Auler et al., 2019). The NDVI helps identify distinct types of vegetation (Jensen, 2009).
Lineament Intensity (LI)	1.23 – 19.63 km/km ²	The genesis of IFCs is often associated with erosive processes that take place during the evolution of slopes, which is facilitated by water infiltration through discontinuities and subsequent chemical and mechanical erosion of the less resistant material usually found below the ferricretes (Dias and Bacellar, 2021). Furthermore, IFCs are structurally controlled (Piló and Auler, 2009; Auler et al., 2022), and these discontinuities represent preferential paths for water inflow. The lineament intensity is the most suitable parameter to represent the surface permeability of the rock mass (Singhal and Gupta, 2010; Florinsky, 2016) with low matrix permeability, such as ferricretes.
Height Above Nearest Drainage (HAND)	0 – 455 m	The genesis of IFCs may be initially associated with the hypogene and saturated environment, which evolves into the vadose phase (Auler et al., 2014; Calux et al., 2019). In general, water flows are not permanent inside IFCs and internal hydrological interconnections are rare (Auler et al., 2014). Due to these characteristics, it can be concluded that IFCs are currently found above the regional water table (Auler et al., 2019). However, the use of HAND aids to map the depth of the temporary saturation

Table 13: Main reasons for choosing the conditioning factors adopted in this study.

FACTORS	MIN – MAX*	REASONS
		zone in the contact between the ferricrete and the itabirite (Auler et al., 2022), forming an interflow that feeds the intermittent drainage channels.

*Minimum and maximum values of each conditioning factor obtained for the study area (Nola and Bacellar, 2021)

5.2 MATERIALS AND METHODS

The methodological steps proposed in this study (Figure 21) can be grouped into six sets of activities: (1) collection and construction of an IFC database and the explanatory factors for the development of IFCs; (2) preparation of training and testing datasets by means of random sampling; (3) application of the RF classification; (4) mapping of the susceptibility to the occurrence of IFCs in a GIS environment; (5) validation of the RF classifier and performance analysis, and (6) evaluation of the importance of each variable.

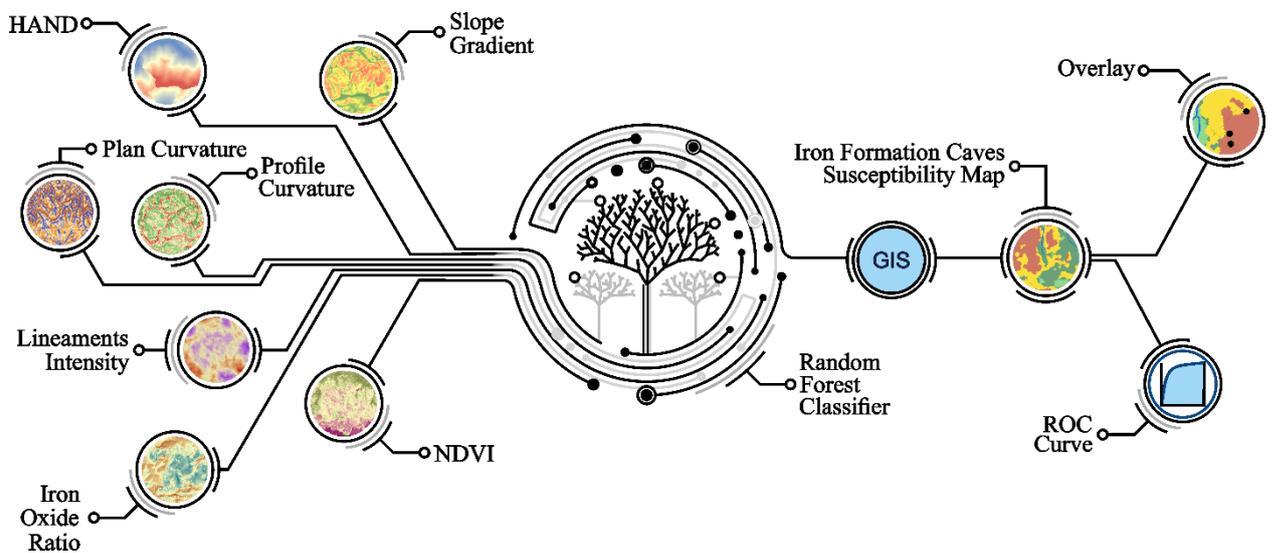


Figure 21: Schematic chart illustrating the methodological steps adopted in this study.

The RF algorithm was implemented on Orange 3.0 (Demsar et al., 2013). This software is open source, and is specific for machine learning, with a great diversity of data mining analysis tools in Python scripting.

The classification result was processed in a GIS environment and was later validated using performance evaluation metrics and inventory overlay for quantitative analysis of the distribution of caves in each class of susceptibility.

5.2.1 IFC Inventory

An IFC inventory was produced from data collected from the Brazilian Speleological Information Registry (CECAV, 2021), to assist the processes of evaluating and validating the results. This is a Brazilian speleological heritage database that gathers the caves prospected by individuals, groups or institutions via scientific publications or direct registration in the system. Fifty-five caves of the study area have already been registered, showing the high density of these features in this type of environment (8 caves/km²). All these caves were included in this study. The spatial representations of the caves were converted to pixel format, with a resolution of 10 m. Pixels representing locations where no caves exist were randomly selected from the non-cave area.

5.2.2 Data Input

The conditioning factors were produced separately using the database listed in Table 14. After the elaboration of the factors database, the products (rasters) were converted into numerical matrices as input to the classification processing. This conversion was performed in ArcGIS (ESRI, 2012).

The construction of an adequate database composed of spatial data is defined by the uniformity of the resolution of the rasters and the projection system. The parity of the geographic positions of the centers of each pixel must be the basis for an accurate combination of rasters, as the numerical matrix of the database of this study is formed from the information of each pixel. In this way, a standard 10 m-resolution raster was created, and all the factors were standardized according to this configuration.

Table 14: Database and geoprocessing environments of the conditioning factors.

FACTORS	DATABASE	GEOPROCESSING ENVIRONMENT
IOR	Produced from bands 2 (490 nm) and 4 (665 nm) captured on 07/13/2016 by Sentinel 2A mission.	ArcGIS (ESRI, 2012)
Slope gradient, Plan curvature, Profile curvature and HAND	Produced from a 10 m-resolution DEM, derived from a cloud of points acquired by LIDAR.	LAStools software, ArcGIS (ESRI, 2012) and TerraView (INPE, 2019)
NDVI	Produced from bands 3 (630 to 690 nm) and 5 (760 to 880 nm) captured by the satellites of the RapidEye mission, with the RapidEye Earth Imaging System (REIS) sensor. The orthorectified scene was supplied by the Minas Gerais State Forest Institute (IEF).	ENVI (ENVI, 2001)
Lineaments intensity	Produced from a set of hillshades with different lighting directions. Generated from the radar image (ALOS/PALSAR) captured by Advanced Land Observing Satellite (ALOS) with the Phased Array type L-band Synthetic Aperture Radar (PALSAR) sensor and obtained free of charge from the Alaska Satellite Facility (ASF) database.	ArcGIS (ESRI, 2012)

5.2.3 Random Forest Classifier

In this study, a binary classifier was used to describe the susceptibility to IFC occurrence using the Random Forest (RF) algorithm. Random Forest was developed by Breiman (2001) as an ensemble method formed by a set of decision trees, each one constructed from a random subset of a training set in order to produce repeated predictions of the same problem. The definitive answer is obtained by a majority vote on the predictions of individual trees, that is, it will be the class selected by the majority of the trees that constitute the forest (Breiman, 2001; Cutler et al., 2012; Biau, 2012).

Breiman (2001) defines that RF is a classifier that uses a set of structured trees $\{h(x, \theta_k), k=1, 2, \dots, J\}$ where J is the number of trees in the forest and $\{\theta_k\}$ are independently identically distributed random vectors. Each tree casts a unit vote for the most popular class at input x .

The RF algorithm creates several trees from a random selection of training data (bootstrap) and the data that is not chosen is called “out-of-bag”. These data are extremely useful for generalization errors and variable importance (Liaw and Wiener, 2002; Cutler et al., 2012).

For a better performance, there are three parameters that can be adjusted: (i) “m”, the number of randomly selected predictor variables chosen at each node; (ii) “J”, the number of trees in the forest, and (iii) “tree size”, as measured by the smallest node size for splitting or the maximum number of terminal nodes. In classification, the value of “m” is defaulted to \sqrt{M} , where “M” is the total number of attributes in the prediction (Breiman, 2001; Cutler et al., 2012). The value of “J” should not be too small, otherwise the out-of-bag estimate can be unstable and inaccurate (Cutler et al., 2012). It is noteworthy that increasing the number of trees does not automatically improve performance, but it can increase the computational load (Taalab et al., 2018). When trees are very large, to adjust a tree depth limit using the “tree size” parameter can be useful to avoid overfitting (Segal and Xiao, 2011; Cutler et al. 2012).

From a computational point of view, RF stands out for the following characteristics (Cutler et al., 2012): speed in training and prediction; classification depends only on one or two tuning parameters; possibility to get the estimate of generalization error; and possibility to measure the importance of each variable for the performance of the method. RF can model complex interactions from continuous or categorical variables, even with missing values, and at the same time it handles irrelevant predictor variables well (Cutler et al., 2012).

A particularly useful feature of RF is its ability to rank predictor variables in order of importance. Several measures, such as Gini Importance, Information Gain, Gain Ratio, Relief-F, are used to study the importance of different variables in RF (Kira and Rendell, 1992; Singh and Gupta, 2014; Quinlan, 2014). The most commonly used method in RF is the Gini Importance (Qi, 2012). The Gini Importance (also Mean Decrease in Impurity) is directly derived from the splitting function called Gini Index to determine which attribute to split during the tree learning phase (Menze et al., 2009; Qi, 2012). In other words, the Gini Index is the way used to decide the ideal division (split) of a root node and subsequent divisions (Menze et al., 2009).

For a single tree, the Gini Importance is defined as the sum of the Gini Index reduction over all nodes in which the specific feature is used to split. For a forest, the Gini Importance is defined

as the average of its importance among all trees in the forest (Cutler et al., 2012). The more the attribute decreases the impurity, the more important the attribute is (Qi, 2012).

5.2.4 Training and Prediction

For classification, training and test data sets, which correspond to the factors that influence the IFC occurrence, and targets, which correspond to the IFC inventory data that is used to classify the training and testing data, are needed in a supervised learning. Thus, RF can predict the possibility of a sample to belong to one of the categories: having a cave (target 1) and not having a cave (target 0).

The database, containing balanced classes, was randomly divided in 70% for training and 30% for testing. The RF algorithm was performed employing a 10-fold cross-validation approach, to reduce the variability of the model results. K-fold cross-validation assumes random division of a dataset into k equal-sized subsamples. The model is trained using k-1 folds, and the k-th fold is used as a test set. This procedure is repeated k times each time a different fold is used as the validation set.

The trees were cultivated without pruning (Breiman, 2001) and the default value of “m”, that is, \sqrt{M} , was used, in this case $\sqrt{7}$, being 7 the number of factors. To adjust the value of the number of trees in the forest (J), values between 1 and 100 were tested, seeking to observe the stabilization of the accuracy value. This stabilization was observed close to 30 (Figure 22).

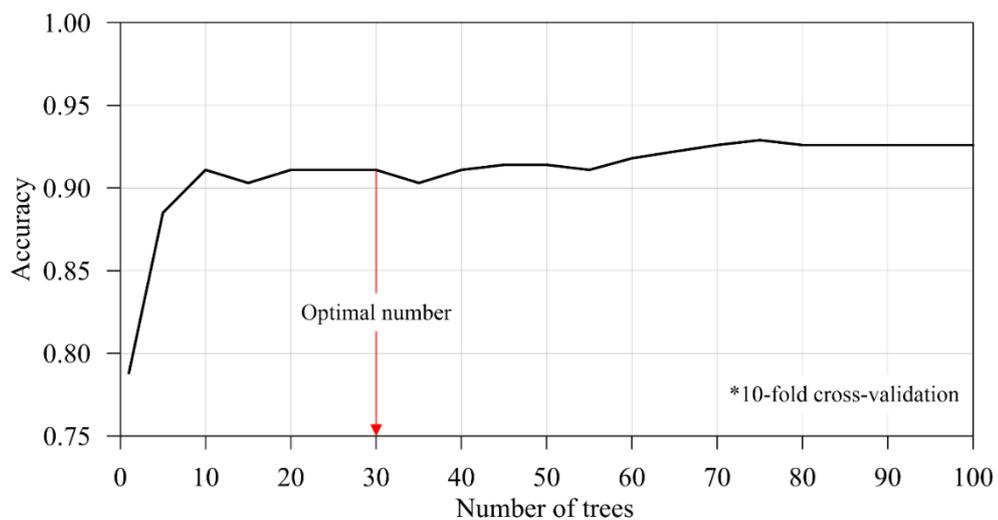


Figure 22: Random Forest Accuracy based on the number of trees.

5.2.5 Validation and Evaluation

The performance of predictive models is usually evaluated with metrics based on the confusion matrix: Accuracy, Precision, Recall and F1, in addition to the ROC Curve analysis (Maria Navin and Pankaja, 2016; Bressan et al., 2020). The confusion matrix presents the real values and the predictive values to visualize the performance of an ML algorithm, where the conditions of True Positive (TP), True Negative (TN), False Positive (FP), and False Negative (FN) are represented (Hossin and Sulaiman, 2015). The metrics: Accuracy (Equation 9) measures the ratio of correct predictions over the total number of evaluated instances; Precision (Equation 10) measures the positive patterns that are correctly predicted from the total predicted patterns in a positive class, Recall (Equation 11) measures the fraction of positive patterns that are correctly classified, and F1 (Equation 12) represents the harmonic mean between recall and precision values (Hossin and Sulaiman, 2015, Tharwat, 2021).

$$Accuracy = (TP + TN)/(P + N) \quad (\text{Eq. 9})$$

$$Precision = TP/(TP + FP) \quad (\text{Eq. 10})$$

$$Recall = TP/(TP + FN) \quad (\text{Eq. 11})$$

$$F1 = (2 * Precision * Recall)/(Precision + Recall) \quad (\text{Eq. 12})$$

The ROC Curve is the most used graphical evaluation method in interpreting classification performance, together with Accuracy (Fawcett, 2006; Tharwat, 2021). The representation of the expected performance is given by the AUC, which is the area under the ROC curve. A realistic classifier must have an AUC value greater than 0.5, reaching 1.0 when the classifier performs well (Fawcett, 2006).

5.3 RESULTS AND DISCUSSIONS

Even though RF is a robust technique and widely used in several research areas, the presence of redundant data can make the importance scores of these variables misleading, and consequently decrease their accuracy (Kubus, 2018; Danasingh et al., 2020). Tolosi and Lengauer (2011) refer to this phenomenon as "correlation bias" and indicate that RF becomes unstable in the presence of high correlations between variables. Strobl et al. (2008) note that correlation bias causes RF to have a preference for correlated predictor variables. This directly

reflects on the importance of the variable, as less relevant variables may replace predictive ones (due to correlation) and thus receive undeserved, boosted importance. Thus, aiming to identify factors that could have this behavior, a correlation analysis was performed between the predictor variables.

Since the data were not normally distributed and it is not known whether the relationship between the variables is linear, the estimation of Spearman’s rank correlation coefficient (also known as Spearman’s rho) was selected. Figure 23 presents the correlation matrix among the seven research variables, where it is possible to observe that the correlations between them are weak. The highest value obtained for the Spearman’s correlation coefficient was 0.45 for IOR and HAND, and -0.45 for NDVI and HAND. This means, for example, that approximately 20% of IOR variability can be explained by its relationship with HAND values, in terms of shared variance (coefficient of determination). In other words, even the factors that showed the highest absolute value of Spearman’s coefficient are not strongly correlated (Mordeson et al., 2022). This result indicates that there is no redundancy among the conditioning factors.

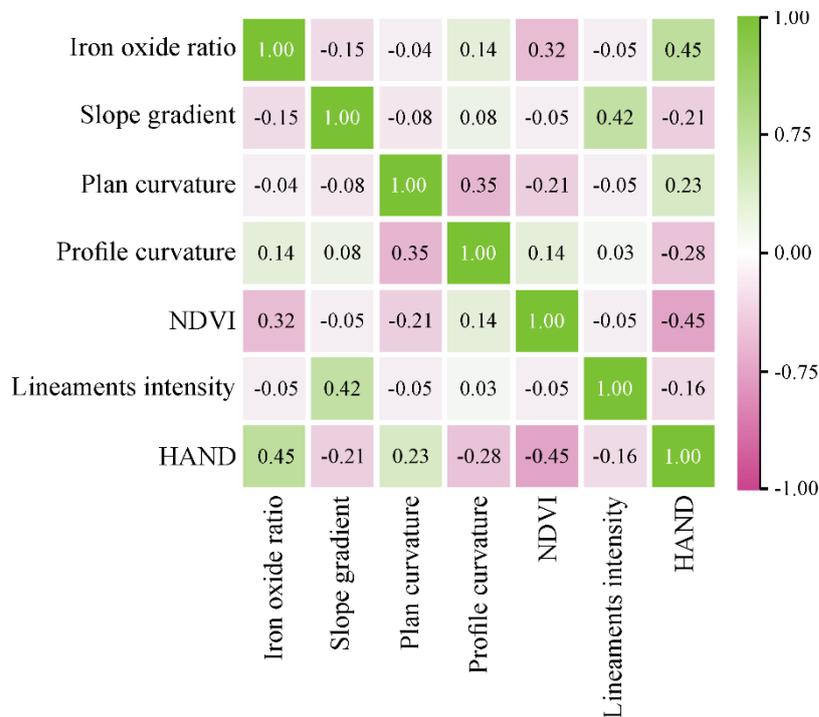


Figure 23: Heatmap representation of the Spearman’s correlation matrix for the seven research variables.

The application of the RF classifier resulted in a susceptibility map of IFC occurrences (Figure 24). As mentioned before, RF can predict the possibility of a sample to belong to one of the

categories: having a cave (target 1) and not having a cave (target 0). Figure 24 was elaborated considering the probability of a sample to belong to the class susceptible to the occurrence of IFCs (target 1). For better spatial representation, the susceptibility map of IFC occurrences was subdivided into four classes: low, moderate, high, very high susceptibility, using the Natural Breaks method. The low and moderate susceptibility classes correspond, respectively, to 20.7% and 19.8% of the study area. The high susceptibility class accounts for 31.6%, while the very high class occupies 27.9% of the study area.

The very high susceptibility class is distributed in the hillslopes located northeast and southwest of the study area. In the center of the study area a more representative and continuous polygon of high susceptibility can be observed between altitudes 1300 and 1600 m. A direct association of the very high susceptibility class with the rupestrian grassland and slope gradients greater than 45% stands out. In the hillslopes where the slope gradient is greater than 45% and the vegetation is denser, the high susceptibility class is predominant. Figure 24 also shows that the high susceptibility class usually surrounds the very high susceptibility class. In turn, the low susceptibility class predominates in areas where slope gradients are less than 30%, mainly at altitudes lower than 1300 m. Low susceptibility is associated with gallery forests or seasonal semideciduous forests. This class is represented as smaller polygons on the plateaus of the NW-SE ridge, where the vegetation is rupestrian grassland. The moderate susceptibility class is scattered throughout the map, between the low and high susceptibility classes, or forms isolated polygons within the low susceptibility class.

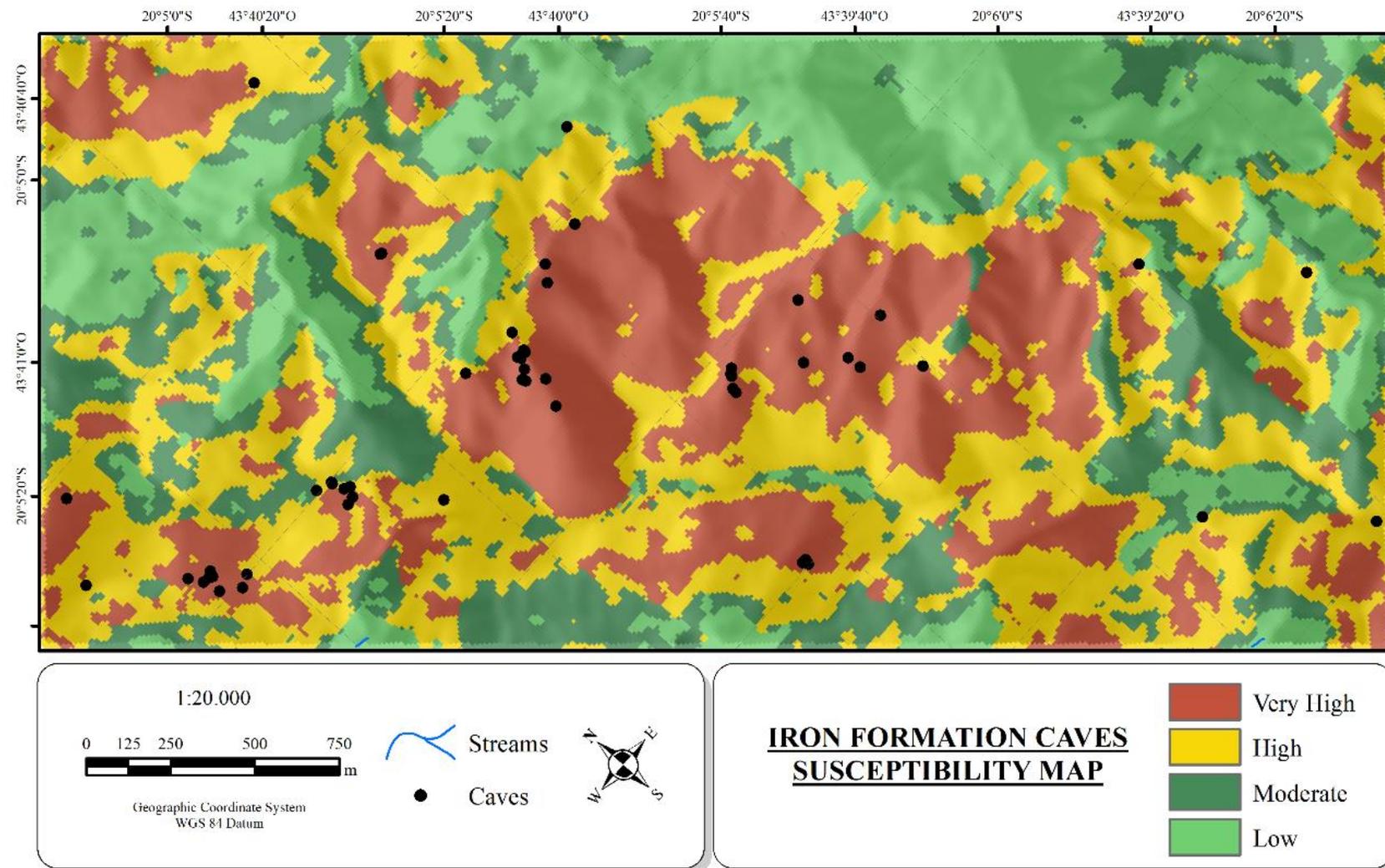


Figure 24: Susceptibility map of IFC occurrences, drawn using the Random Forest Classifier.

Using the training and validation sets, the success and prediction rates were obtained by plotting the ROC curve with a limit of 100 analyses (Figure 25). The values of the areas under the ROC curve (AUC values) were 0.954 and 0.957 for the success and precision rates, respectively. These values are close to 1.0, indicating a very good performance of the predictive model, and their similarity indicates that the RF method is reliable (Rasyid et al., 2016).

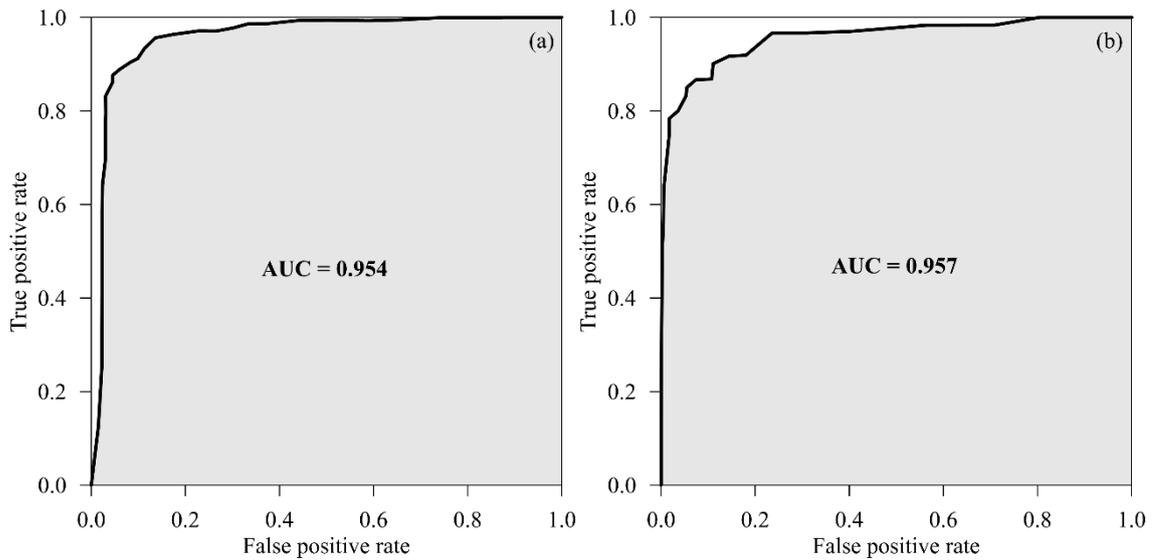


Figure 25: ROC Curves: (a) success rate curve using the training dataset; and (b) predicted rate curve using the validating dataset.

As explained before, metrics based on the confusion matrix were used to evaluate the performance of the predictive method. The results shown in Figure 26 attest that the application of RF has a good predictive capacity in assessing the susceptibility to the occurrence of IFCs, with values greater than 0.9 in training (red bars) and greater than 0.89 in testing (grey bars).

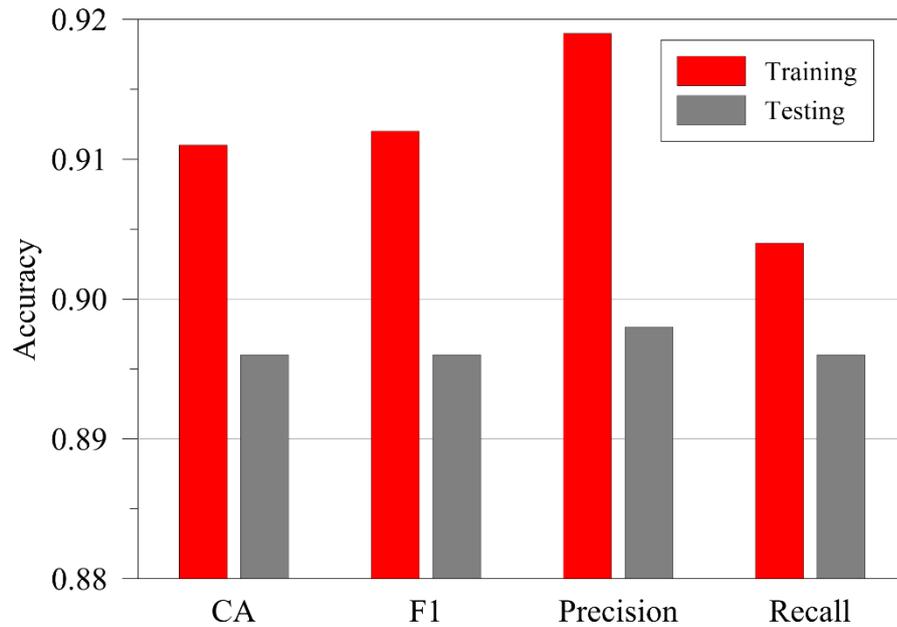


Figure 26: Performance evaluation metrics.

Gini Importance was calculated for the seven factors that contributed to classify the susceptibility to IFC occurrence. According to the results (Figure 27), NDVI yielded the highest Gini value (0.137), followed by HAND (Gini = 0.123), suggesting that these were the factors that most contributed to the classification model. Profile Curvature and Plan Curvature presented the lowest Gini values, respectively 0.013 and 0.005, showing little or no contribution to the prediction process.

It is worth noting that the AHP previously employed by Nola and Bacellar (2021) and the Random Forest approach used in this study present different methods of analysis, which makes it difficult to directly compare these techniques in terms of the importance of each variable in the process. In AHP, a group of specialists choose the most important variables to model the process and attribute to these variables a value from the scale of importance of Saaty (2001) based on a paired comparison. Thus, the variable with the greatest weight is the one that the specialist considers to be the most important, without prior analysis of the IFC dataset. In Random Forest, the algorithm analyzes all the factors within the IFC dataset, and searches among all them the one(s) that yielded the smallest impurity (Gini Impurity) when splitting a node, which corresponds to the highest Gini Importance value(s). Thus, the variable that was most effective in the division between classes “having a cave” (target 1) and “not having a cave” (target 0), is the variable that has the highest Gini Importance value.

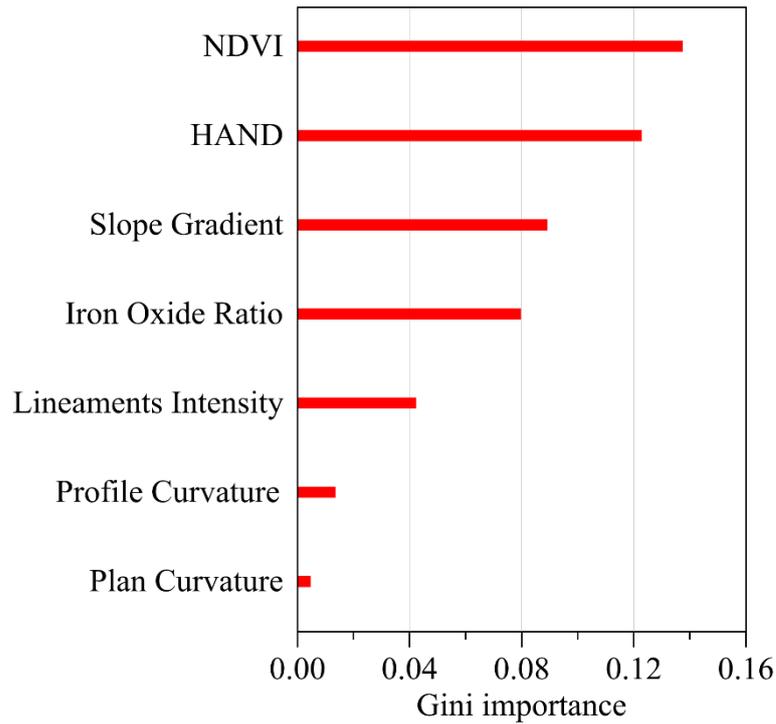


Figure 27: Variable importance expressed in terms of Mean Decrease in Gini Impurity.

The factors that the AHP experts judged to be the most important: Iron Oxide Ratio and Slope Gradient, as detailed by Nola and Bacellar (2021), yielded average Gini Importance values in the present study. However, HAND, judged by the specialists as the least important (Nola and Bacellar, 2021), was one of the variable that best defined the division between the nodes in the RF classification.

In order to assess the correlation between each of the susceptibility classes and IFCs, frequency ratios (FR) were calculated. According to Lee et al. (2012), if FR is greater than 1, the correlation is strong; on the contrary, if FR is less than 1, the correlation between the analyzed data is low. For the low, moderate, high and very high susceptibility classes, the correspondent FR values are respectively 0.09, 0.18, 0.75, and 2.54. It is observed that the very high susceptibility class has the highest FR value, followed by the high susceptibility class, whose FR value is close to 1 (strong correlation).

When comparing the FR values obtained by analyzing the susceptibility maps produced by each technique (Fuzzy Logic and AHP – Nola and Bacellar, 2021, and Random Forest – this study, Figure 28), it is observed that the FR value obtained for the very high susceptibility class was higher for RF than for AHP. This suggests that the RF technique produced a more accurate map

compared to that produced by AHP, as the relationship of the very high susceptibility class with the IFC occurrence is strong (FR = 2.54), stronger than that yielded by AHP.

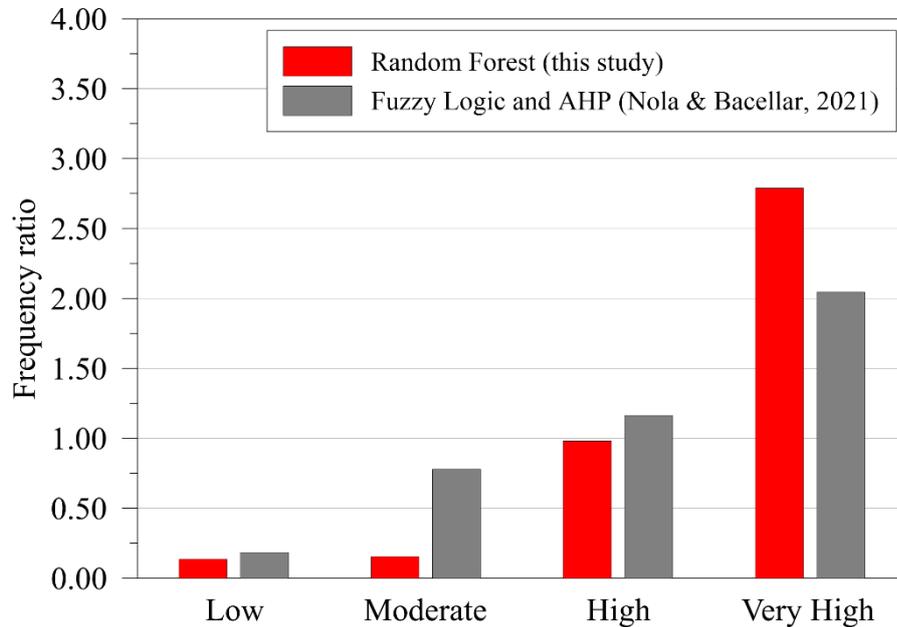


Figure 28: Frequency Ratio values for each susceptibility class generated by Fuzzy Logic and AHP (Nola and Bacellar, 2021), and by Random Forest (this study).

5.4 CONCLUSIONS

This study demonstrates that it is possible to use Machine Learning, adopting the Random Forest algorithm to produce a susceptibility map of IFC occurrences with high accuracy and high prediction rates.

The results of the RF application presented better accuracy and higher Frequency Ratio for the very high susceptibility class, when compared to those obtained by Nola and Bacellar (2021) via Fuzzy Logic and AHP. In addition to the ease of application and interpretation, RF allows to establish which variables are most important in the classification, removing the subjectivity inherent to the methods that depend on the judgment of experts, such as AHP. Therefore, RF has proven to be a more effective technique to studying susceptibility to the occurrence of IFCs.

Gini Importance showed which were the most relevant variables in RF classification. Among the seven research factors, NDVI and HAND were the ones that most contributed to the susceptibility prediction. We suggest the Plan Curvature variable be removed from future studies, as it contributed very little to the susceptibility prediction. As the modus operandi of

the conditioning factors addressed in this study are poorly known, assessing the relevance of each factor may contribute to the understanding of the genesis of IFCs.

For supervised learning using the RF classification technique, it is essential to define the targets, that is, to identify places where caves are existent and inexistent. The challenges involved in this study include: (i) a better way of sampling the areas where the caves are inexistent, to control and refine the most effective values in the classification of target 0 (no caves), and (ii) efforts to expand the training database in order to improve the mapping of the high and very high susceptibility areas.

We expect that the methodological steps detailed in this study will encourage planners and decision makers from mining ventures and environmental agencies to adopt them with the purpose of mapping the susceptibility to the occurrence of IFCs totally on the basis of remote data, making speleological prospecting faster, cheaper and safer.

ACKNOWLEDGEMENTS

The authors would like to thank CAPES -Coordination for the Improvement of Higher Education Personnel, FAPEMIG – Minas Gerais Research Funding Foundation, CNPq - National Council for Scientific and Technological Development, and UFOP- Federal University of Ouro Preto for academic and scientific support. The Geotechnical Group (NUGEO) of the Escola de Minas - UFOP offered the laboratory facilities. The authors are also grateful to Vale SA for the opportunity to do this work and for providing data from the study region.

CAPÍTULO 6

REDES NEURAIAS ARTIFICIAIS APLICADAS AO MAPEAMENTO DA SUSCETIBILIDADE À OCORRÊNCIA DE CAVIDADES FERRÍFERAS

Os resultados e as discussões obtidos neste trabalho de doutorado estão aqui apresentados no formato de artigo científico, que ainda não foi submetido a nenhum periódico³,

ABSTRACT:

O principal objetivo deste estudo é avaliar o desempenho de redes neurais artificiais (ANN) na suscetibilidade a ocorrência de cavidades ferríferas na região da Serra do Gandarela, Quadrilátero Ferrífero (MG) e comparar os resultados com aqueles obtidos com as técnicas Floresta Aleatória (*Random Forest* - RF) e Processo Analítico Hierárquico (*Analytic Hierarchy Process* - AHP). Para essa comparação, foi utilizado o banco de dados de inventário de cavidades e o mesmo conjunto de fatores condicionantes para seu desenvolvimento. Para a aplicação da ANN, selecionou-se a rede *Perceptron* multicamadas e o algoritmo de aprendizado *backpropagation*. Várias arquiteturas foram treinadas até ser selecionada a rede com o melhor desempenho. Os valores de acurácia para a rede escolhida foram 0,833, para treinamento, e 0,783, para teste. Os valores de AUC foram 0,886, para treinamento, e 0,866, para teste. Ao final, confeccionou-se um mapa de suscetibilidade a ocorrência de cavidades ferríferas. Os resultados indicam que o modelo aplicado não possui capacidade de generalização, pois impossibilitou classificar dados nunca vistos, apesar de seu desempenho ter sido considerado satisfatório durante o treinamento. Quando os resultados das três técnicas são comparados, observa-se que Floresta Aleatória apresentou a melhor capacidade preditiva quando utilizados os parâmetros escolhidos para esta aplicação.

KEYWORDS: Iron Formation Caves, Susceptibility Map, Machine Learning, Artificial Neural Network, Brazil

³ Artigo não submetido.

³ Autores: Iraydes Tálita de Sena Nola^a; Luis de Almeida Prado Bacellar^a

^a Núcleo de Geotecnia, Escola de Minas, Universidade Federal de Ouro Preto, MG - Brazil

6.1 INTRODUÇÃO

Cavidades ferríferas são feições cársticas que se desenvolvem em materiais ricos em ferro, normalmente em zonas topicais (Simmons, 1963). No Brasil, elas são encontradas principalmente no Quadrilátero Ferrífero e na Província Mineral de Carajás. Atualmente, as prospecções espeleológicas nestas localidades se expandem, devido a estreita relação entre as cavidades ferríferas e os empreendimentos minerários de exploração de minério de ferro. Essa demanda, associada as exigências ambientais relativas ao planejamento da exploração mineral e ao licenciamento ambiental, suscitam a necessidade do desenvolvimento de técnicas de mapeamento remoto capazes de otimizar a prospecção espeleológica.

Recentemente, Nola e Bacellar (2021, 2022) propuseram a aplicação de técnicas de multicritério por Processo Hierárquico Analítico (AHP) e Lógica Fuzzy, e também a aplicação de aprendizado de máquina por meio da técnica Floresta Aleatória (RF), pra mapear a suscetibilidade de ocorrência de cavidades ferríferas. Ambas as técnicas apresentaram boa capacidade preditiva à ocorrência de cavidades ferríferas, se tornando ferramentas eficientes na tomada de decisão.

Além de RF, uma das técnicas de aprendizado de máquina que se destaca por seu poder computacional, por sua habilidade de aprender e, conseqüentemente, generalizar, são as Redes Neurais Artificiais (ANN) (Haykin, 2004). As ANNs são capazes de detectar relações não lineares complexas entre as variáveis, apresentam independência da distribuição estatística dos dados e a possibilidade de utilizar variáveis tanto quantitativas quanto qualitativas (Bragagnolo et al., 2020).

Exemplos recentes que ilustram a aplicação de ANN em geociências, especialmente aplicadas ao mapeamento, incluem avaliação da suscetibilidade a: movimentos de massa (Ortiz e Martínez-Graña, 2018; Aditian et al., 2018; Mandal e Mondal, 2019; Shahri et al., 2019; Bragagnolo et al., 2020), terremotos (Xie et al., 2018; Tian et al., 2019; Rajabi et al., 2021); a processos erosivos (Sarkar e Mishra, 2018; Band et al., 2020; Roy e Saha, 2021); e inundações (Falah et al., 2019; Wang et al., 2020; Kalantar et al., 2021). No caso de ambientes cársticos, a ANN tem sido aplicada para modelar a hidrologia de bacias cársticas (Trichakis et al., 2011; Kovačević et al., 2018; Sezen et al., 2019), para classificar as feições de relevo (Sepahvand et al., 2019) e para prever o tamanho das cavidades (Liu et al., 2018).

Para avaliar a suscetibilidade a ocorrência cavidades ferríferas, Nola e Bacellar (2021) selecionaram uma área na Serra do Gandarela, no IQ, onde existe uma alta concentração de cavidades ferríferas (8 cavidades/km²). Os autores construíram um banco de dados com sete fatores relativos à gênese e principais características das cavidades ferríferas, além da localização das cavidades existentes na região. No presente trabalho, pretende-se utilizar este mesmo banco de dados para avaliar a aplicação de ANN no mapeamento da suscetibilidade à ocorrência de cavidades ferríferas, e comparar os resultados àqueles obtidos com as técnicas AHP e RF.

6.1.1 Área de Estudo

A área de pesquisa possui 7,2 km² e está localizada na porção sudoeste da Serra do Gandarela, na divisa entre as cidades de Rio Acima e Santa Bárbara, no estado de Minas Gerais (Figura 29). A Serra do Gandarela pertence ao conjunto de serras formadas por processos erosivos diferenciais (Salgado et al., 2004) que se destacam no contexto geológico-geomorfológico do IQ por serem constituídas por itabiritos ou quartzitos e por apresentarem as maiores altimetrias dentro da província mineral (Varajão, 1991), chegando a cerca de 2000 m de altitude (Serra do Caraça).

Na região, além dos itabiritos e quartzitos, podem ser encontrados dolomitos, itabiritos dolomíticos, xistos e filitos, além de materiais supergênicos desenvolvidos principalmente sobre itabiritos, denominados localmente cangas (*ferricretes*) (Alkmim e Marshak, 1998). O substrato rochoso controla diretamente o relevo, que por sua vez está intimamente ligado à vegetação (Salgado et al., 2004; Lamounier et al., 2010). Assim, a porção mais alta da área (1660 -1600 m), correspondente aos platôs da Serra do Gandarela, é constituída por itabiritos cobertos por cangas, onde predomina a vegetação do tipo campos rupestres ferruginosos (Fernandes, 2016; Marent e Portilho, 2017) (Figura 29). A encosta sudeste da serra é representada por rampas formadas por cangas, entre 1600 e 1450 m, e por dolomitos, abaixo de 1450 m de altitude. Sobre as cangas observam-se campos rupestres ferruginosos associados a campos sujos (Marent e Portilho, 2017). Verifica-se ainda expressivas quebras de relevo no limite entre materiais mais resistentes (cangas sobre itabiritos) e menos resistentes (dolomitos) que coincidem com uma mudança abrupta da vegetação para mata de galeria ou floresta estacional semidecídua, representante do bioma da Mata Atlântica (Oliveira et al., 2005; Lamounier et al., 2010; Marent e Portilho, 2017) (Figura 29).

Cavidades naturais podem ser encontradas na Serra do Gandarela em dolomitos, quartzitos, formações ferríferas bandadas e em cangas. Na área de estudo foram registradas 55 cavidades ferríferas (CECAV, 2021), o que equivale a 7,6 cavidades/km², mostrando a alta potencialidade de ocorrência dessas feições cársticas não só nesta como em outras serras do Quadrilátero Ferrífero.

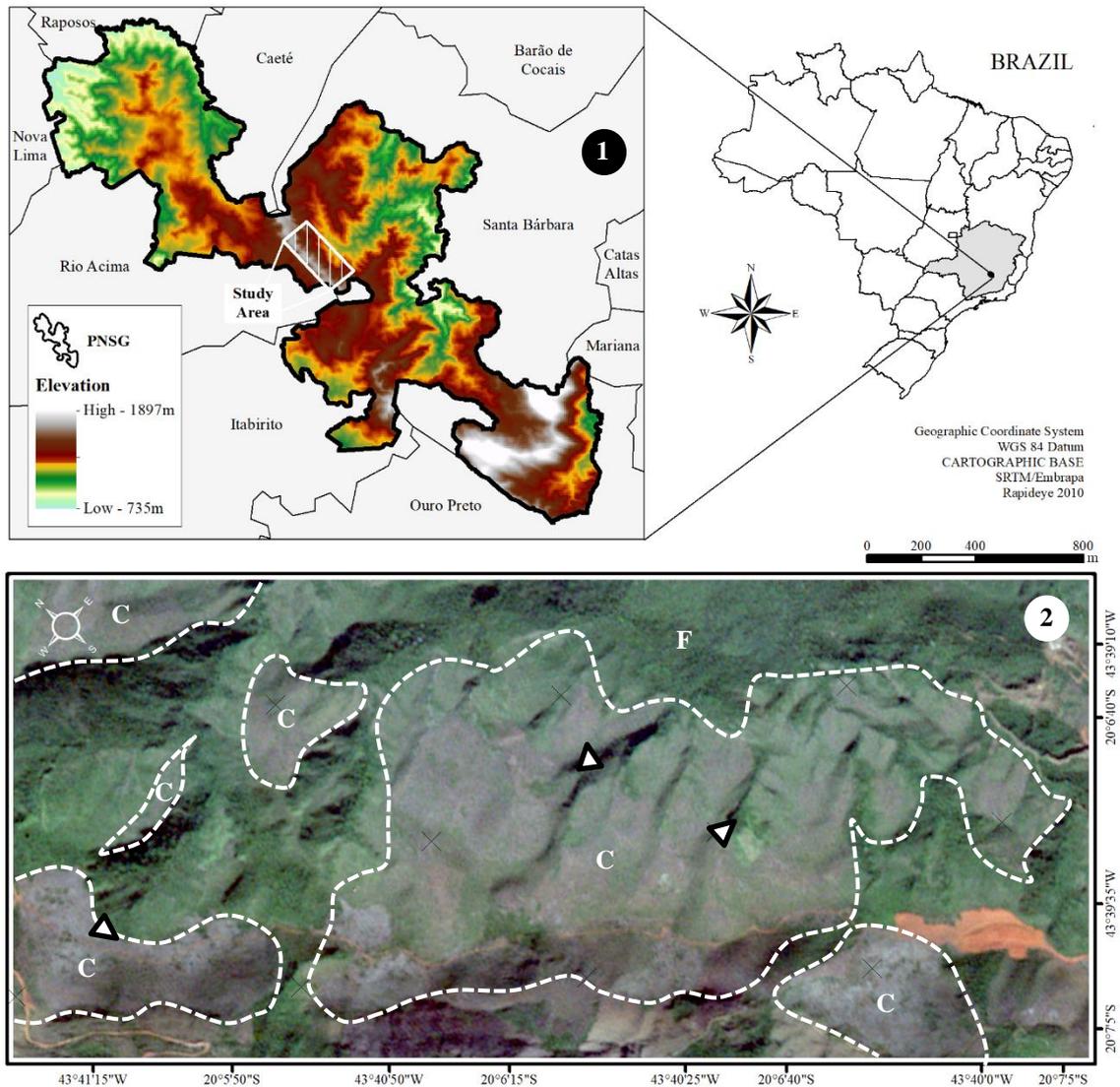


Figura 29: 1) Localização do PNSG em um contexto regional com destaque para o polígono representativo da área de estudo. 2) Imagem de satélite da área de estudo onde se observa o relevo, a vegetação e a intervenção antrópica (solo exposto de coloração laranja). Nessa figura a letra C indica a região onde predomina a vegetação de campos rupestres ferruginosos, a letra F indica floresta estacional semidecídua, e os triângulos representam algumas quebras de relevo na área.

Desde 2014, a preservação ambiental da região se dá por um decreto federal que criou o Parque Nacional Serra do Gandarela (PNSG) (Brasil, 2014). Este parque foi criado com o objetivo de proteger e conservar a diversidade biológica, geológica, hidrológica, espeleológica, além do conjunto geomorfológico composto por serras e platôs associados às cangas, incluindo campos rupestres e remanescentes da mata atlântica.

Mesmo havendo a proteção da região por lei, localmente, podem ser vistas interferências antrópicas relacionadas às áreas de pastagens, nas posições mais baixas do relevo, e às áreas de mineração em posições mais elevadas, associadas à mineração de ferro, de bauxita e de canga (Lamounier et al., 2010). Algumas intervenções associadas à prospecção mineral, como o solo exposto proveniente da abertura de trincheiras e de realização de sondagens, anteriores ao decreto que cria o PNSG, podem ser observadas na Figura 29.

6.1.2 Cavernas Ferríferas

As cavernas ferríferas encontram-se associadas às cangas, formações ferríferas bandadas e minério de ferro (Simmons, 1963; Piló et al., 2015a; Dutra, 2013), especialmente naquelas variedades com maior teor relativo de ferro (Calux, 2013). Geralmente estão posicionadas nas bordas das coberturas da canga em razão da existência de quebras de relevo resultantes do recuo erosivo (Piló e Auler, 2009; Auler et al., 2019; Dias e Bacellar, 2021). O processo de formação pode acontecer a partir da dissolução e/ou lixiviação do ferro (Pinheiro e Maurity, 1988), seguido por erosões mecânicas devido a fluxos subsuperficiais (Auler e Piló, 2005). A vegetação na área de ocorrência das cavernas ferríferas é predominantemente campo rupestre ferruginoso (Stávale, 2012; Piló et al., 2015a), mas as entradas das cavernas são estreitas e cobertas por vegetação de maior porte (Simmons 1963; Dutra, 2013).

A gênese das cavernas ferríferas está frequentemente associada a processos erosivos das encostas, o que é facilitado pela infiltração de água por discontinuidades e a subsequente erosão química e mecânica do material menos resistente, geralmente encontrado abaixo das cangas (Auler et al., 2019; Dias e Bacellar, 2021). Desenvolvem-se nas bordas de platôs, incluindo baixa, média e alta vertente, associadas as quebras de relevo e, também ao redor de drenagens e lagoas (Simmons, 1963; Piló e Auler, 2009). Geralmente não possuem fluxos de água permanentes em seu interior e a interconexão hidrológica interna é rara (Auler et al., 2014), o que leva a compreensão de que as cavernas ferríferas são encontradas acima do lençol freático

regional (Auler et al., 2019); portanto, a sua evolução está associada a um ambiente vadoso, embora seu início tenda a ocorrer em ambiente freático ou transicional (Calux, 2013).

Cavidades ferríferas apresentam dimensões reduzidas e situam-se próximas à superfície, desenvolvendo-se concordantemente ao relevo, com inclinação paralela à vertente (Dutra, 2017), resultando numa espessura de teto muito pequena, inferior a 5 m (Auler et al., 2019). Suas câmaras são irregulares e/ou de padrão linear, este associado as feições litoestruturais (Piló e Auler, 2009). Elas apresentam controle estrutural evidenciado por descontinuidades limitando as paredes e os tetos (Piló et al., 2015b) e marcado pelas passagens estreitas entre câmaras, diversas direções de condutos, semi-paralelos ou paralelos (Dutra, 2013).

6.2 ESTUDOS ANTERIORES

O sistema cárstico é estudado há décadas, principalmente as feições desenvolvidas em rochas carbonáticas e seus processos de formação. Apesar de terem sido identificadas nos meados do século 19 (Auler et al., 2022), as cavidades ferríferas só começaram a ser sistematicamente estudadas nas últimas décadas (Auler e Piló, 2005; Piló e Auler, 2009; Auler et al., 2019; Auler et al., 2022). Contudo, questões relevantes para a prospecção e mapeamento e mapeamento das cavidades ferríferas ainda necessitam de estudos.

Diante deste cenário, os autores desta pesquisa vêm estudando técnicas a fim de mapear a suscetibilidade de ocorrências de cavidades ferríferas no Quadrilátero Ferrífero, Brasil. Uma das técnicas aplicadas foi a análise de multicritério por meio de AHP com padronização de variáveis por lógica fuzzy (Nola e Bacellar, 2021). Neste trabalho foram selecionados alguns fatores que ajudariam a explicar a ocorrência de cavidades ferríferas. Estes fatores foram escolhidos levando em consideração as características morfológicas, de gênese e evolução das cavidades ferríferas.

Nola e Bacellar (2021) então sugeriram sete fatores, são eles: razão de óxido de ferro; declividade; *Normalized Difference Vegetation Index* (NDVI), curvatura horizontal; curvatura vertical; intensidade de lineamentos estruturais; e *Height Above Nearest Drainage* (HAND). Por meio da razão de óxido de ferro buscou-se mapear os materiais geológicos superficiais ricos em ferro (Ducart et al., 2016). A declividade está relacionada aos processos de erosão superficiais e subsuperficiais, assim como a curvatura horizontal e a curvatura vertical, sendo úteis na avaliação da velocidade do fluxo d'água, do sentido do fluxo, e do transporte de

sedimentos (Wilson e Gallant, 2000; Wysocki et al., 2011; Summerfield, 2014). A partir do cálculo do NDVI procurou-se identificar os diferentes tipos de vegetação (Jensen, 2009). A intensidade de lineamentos foi utilizada para avaliar a infiltração de água nos litotipos da área, que se comportam como maciços fraturados (Singhal e Gupta, 2010; Florinsky, 2016). Por fim, a utilização do HAND visou mapear a profundidade da zona de saturação temporária no contato entre a canga e o itabirito (Auler et al., 2022), formando um interfluxo que alimenta os canais de drenagem intermitentes (Dias e Bacellar, 2021).

Os resultados apresentados por Nola e Bacellar (2021) demonstram que a técnica de AHP mostrou-se adequada para o mapeamento de áreas suscetíveis ao desenvolvimento de cavidades ferríferas. Os fatores razão de óxido de ferro, declividade e NDVI –obtiveram o maior peso entre os julgamentos dos especialistas, 0,367 e 0,257 e 0,202, respectivamente. A taxa de predição apresentada foi de 0,852, que indica um desempenho bom e confiável da técnica. Além disso, a relação de frequência (*frequency ratio* – FR), que indica a correlação entre as cavidades ferríferas e a classe de suscetibilidade, mostrou valores de 2,04 e 1,17 para as classes muito alta e alta, respectivamente, indicando uma forte correlação.

Apesar desta técnica ser amplamente empregada e alcançar bons resultados, AHP utiliza o julgamento de especialistas para obtenção dos pesos dos critérios, partido de uma avaliação inicial puramente subjetiva. Assim, buscou-se então aplicar uma nova técnica com o objetivo de remover do processo a subjetividade do julgamento dos especialistas. Os mesmos autores propuseram a utilização do Aprendizado de Máquina (*Machine Learning* – ML) (Mohri et al., 2018) por meio da técnica RF (Breiman, 2001). O RF é um algoritmo de fácil aplicação e interpretação, utilizado comumente em geociências (Zhu e Pierskalla, 2016; Elmahdy et al., 2020; Tanyu et al., 2021).

A aplicação de RF demonstrou ser eficaz no mapeamento de áreas suscetíveis ao desenvolvimento de cavidades ferríferas (Nola e Bacellar, 2022), com alta taxa de predição, de 0,957. Esta análise foi feita com os mesmos fatores adotados por Nola e Bacellar (2021) e aqueles que mais contribuíram para a análise foram o NDVI, HAND e Declividade, que apresentaram os valores de *Gini Importance* de 0,137, 0,123 e 0,89, respectivamente. Os valores da relação de frequência para as classes muito alta e alta suscetibilidade mostrou uma forte correlação entre as classes e a ocorrência de cavidades ferríferas, apresentando valores de 2,54 e 0,75 respectivamente.

Como observado por Nola e Bacellar (2022), a técnica de RF apresentou melhores resultados de taxa de predição e relação de frequência, para a partição adotada, quando comparados aos resultados produzidos pela aplicação de AHP. Além disso, RF é uma técnica de ML que recebe informações de entrada, aprende, interpreta e classifica novos resultados a partir de um algoritmo que possui apenas três parâmetros de performance, ou seja, não há nenhuma interferência direta do especialista na decisão da relevância dos fatores de predição além da escolha dos parâmetros ideais de treinamento.

6.3 MATERIAIS E MÉTODOS

A ANN é um sistema computacional estruturado para receber informações, interpretá-las e tomar decisões apropriadas a partir de um treinamento (Haykin, 2004). Dentre suas características mais relevantes, destaca-se: (1) a adaptação por experiência envolvida nesta técnica leva os parâmetros a serem ajustados a partir de uma apresentação sucessiva de exemplos relacionados diretamente ao processo a ser descoberto (Silva et al., 2010) e, (2) a partir da habilidade de generalização, a ANN pode ser capaz de, após o treinamento, estimar soluções desconhecidas apenas por meio do conhecimento adquirido (Haykin, 2004; Bragagnolo et al., 2020).

A ANN foi desenvolvida com base em uma rede de múltiplas camadas alimentadas adiante (*feedforward*) (Figura 30). Esta rede consiste de um conjunto de neurônios artificiais que formam a camada de entrada, seguido de uma ou duas camadas intermediárias, e uma camada de saída (Haykin, 2004). Atualmente, essa é a rede mais amplamente utilizada devido a sua capacidade de resolver problemas complexos a partir de uma arquitetura simples (Haykin, 2004).

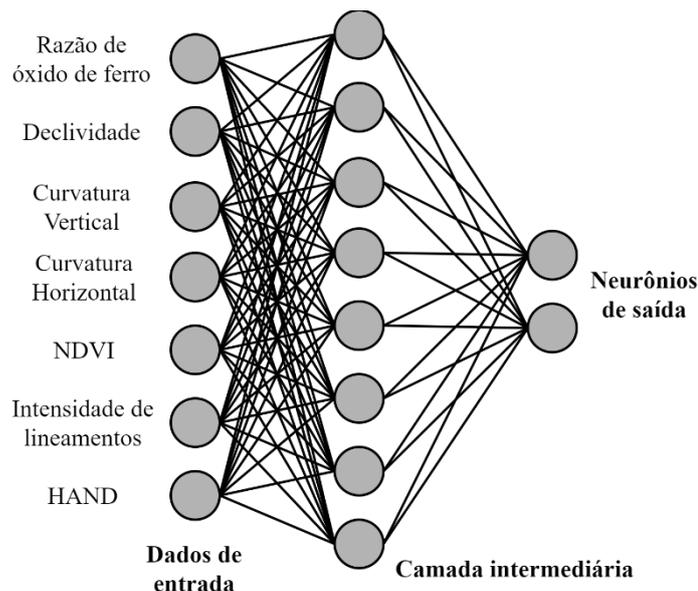


Figura 30: Arquitetura da MLP utilizada.

Para treinamento e posterior predição de novos dados foi utilizado o *software* Orange 3.0 (Demsar et al., 2013). Este software possui código aberto para implementação de diversas técnicas de aprendizado de máquina, supervisionadas ou não supervisionadas, com uma grande diversidade de ferramentas em linguagem de programação Python.

Os dados de entrada da rede constituem o conjunto de fatores selecionados para descrever o processo a ser descoberto. Neste caso, foi utilizado o mesmo conjunto de dados criado por Nola e Bacellar (2021), com os sete fatores empregados previamente (Figura 2). Após o treinamento e a validação da rede, a saída resulta em uma medida de suscetibilidade a ocorrência de cavidades ferríferas um determinado pixel, em uma escala que varia de 0 (baixa suscetibilidade) a 1 (alta suscetibilidade), que correspondem aos dados para elaboração do mapa (Figura 31).

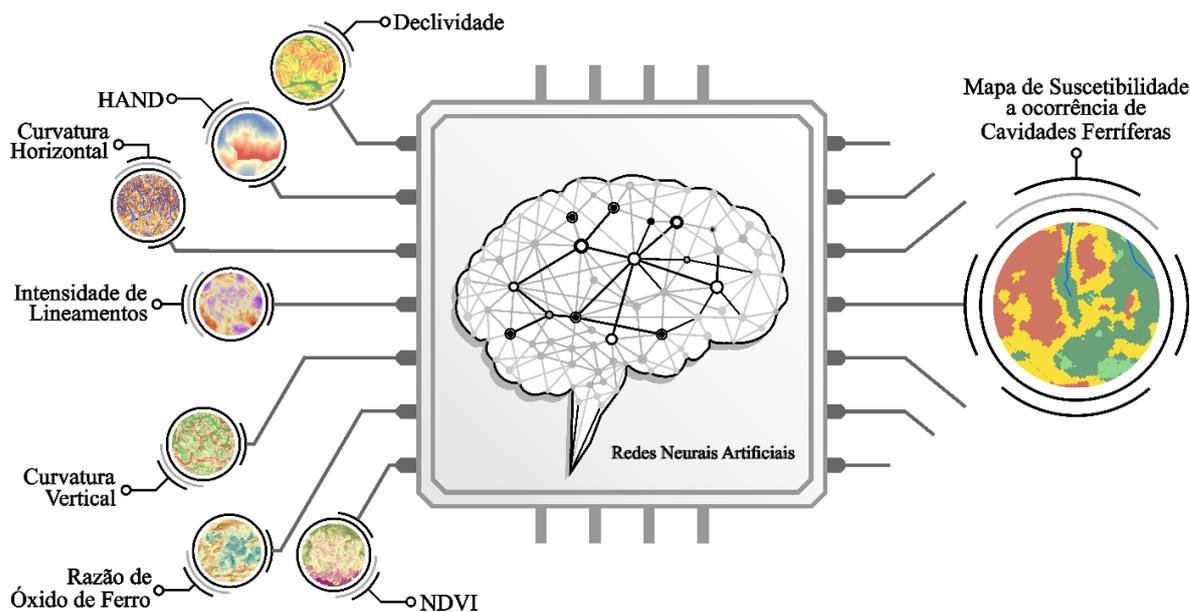


Figura 31: Representação esquemática da aplicação de Redes Neurais Artificiais neste estudo.

6.3.1 Ajuste dos parâmetros e treinamento

O treinamento de forma supervisionada foi realizado pelo algoritmo *backpropagation* (Haykin, 2004), que pode resolver problemas não-linearmente separáveis por meio da rede neural *Perceptron* multicamadas (Silva et al., 2010). Esse treinamento desenvolve-se em três fases (Haykin., 2004; Silva et al., 2016). Na primeira fase o conjunto de entrada é aplicado aos neurônios e o seu efeito se propaga camada por camada até chegar na camada de saída, e o resultado final ser produzido. Na segunda fase, a resposta produzida pelo processamento é comparada à resposta desejada e o erro é calculado. Na terceira fase, o erro é propagado de volta e os pesos sinápticos são ajustados até chegar novamente na camada de entrada. Em síntese, esse ajuste consiste em alterar os pesos sinápticos de forma com que a resposta do algoritmo se mova para mais perto da resposta desejada.

Os parâmetros relativos ao treinamento foram ajustados visando otimizar o uso dos recursos para que a performance de trabalho seja a mais eficiente possível. Este processo é denominado *tunning*. O *tunning* depende não só dos parâmetros, mas também do gasto computacional e do tempo de resposta no processamento dos dados (Haykin, 2004). O procedimento de parada antecipada foi aplicado durante o treinamento para evitar *overfitting* que consiste no ajuste excessivo da ANN aos dados de treinamento (Haykin, 2004). Para o critério de parada foi definido o número máximo de épocas de 200 e a taxa de aprendizado como $\alpha = 2$.

Para a definição do limite máximo de neurônios da camada intermediária, foi aplicada a equação 13 proposta por Netch-Nielsen (1987)

$$N_H \leq 2 N_I + 1 \quad \text{Eq (13)}$$

onde N_H é o número de neurônios na camada intermediária e N_I é o número de fatores de entrada. Considerando 7 fatores de entrada, o número máximo de neurônios na camada adotada foi de 15.

Inicialmente, duas funções de ativação foram utilizadas na avaliação do desempenho da rede neural, a função logística e a função linear retificada (ReLU - *Rectified Linear Unit*). A função logística é especialmente utilizada em modelos que necessitam prever um valor de probabilidade como saída, isto porque o intervalo desta função encontra-se dentro do intervalo de existência de uma probabilidade, entre 0 e 1 (Cho, 2009; Sevgen et al., 2019). A função ReLu é atualmente a função mais utilizada em ANN, por possuir uma derivada de primeira ordem constante, diferentemente da função logística (sigmoideal) que devido à sua curvatura possui uma pequena faixa de derivadas (Rasamoelina et al., 2020).

Após a escolha da função de ativação e do número de neurônios da camada intermediária, a rede neural foi avaliada com base no método *random sampling* (Koch et al., 2013), no qual diferentes conjuntos de dados de treinamento e validação foram gerados aleatoriamente, a fim de analisar a variação do desempenho da aprendizagem. O conjunto de dados de treinamento foram amostrados 30 vezes.

Do conjunto de dados existente para treinamento, ou seja, dados que possuem o valor esperado, 70% foi selecionado randomicamente para o treinamento da rede e 30% para a teste. Os dados são balanceados e foram normalizados entre valores de 0 a 1. Devido ao pequeno tamanho do conjunto de dados de treinamento (384 instâncias), a melhor forma de avaliar a qualidade do modelo é por meio da validação cruzada (Silva et al., 2016). Nesta validação, o conjunto de treinamento é dividido aleatoriamente em K amostras de tamanhos iguais. O modelo é treinado utilizando K-1 amostras e a amostra restante é utilizada como conjunto de teste.

6.3.2 Avaliação do desempenho da ANN

Para medir o desempenho da rede neural treinada foi calculada a medida de acurácia, precisão, recall e F1, usando a matriz de confusão, por meio das equações abaixo:

$$Acurácia = \frac{TP+TN}{I} \quad \text{Eq (14)}$$

$$Precisão = \frac{TP}{TP+FP} \quad \text{Eq (15)}$$

$$Recall = \frac{TP}{TP+FN} \quad \text{Eq (16)}$$

$$F1 = \frac{2*Precisão*Recall}{Precisão+Recall} \quad \text{Eq (17)}$$

Onde TP é são os verdadeiros positivos, TN são os verdadeiros negativos e, I é o número total de instâncias avaliadas. Adicionalmente, foram obtidos os valores de AUC para todos os modelos analisados para melhor interpretação do desempenho da classificação (Hossin e Sulaiman, 2015; Fawcett, 2006). O valor de AUC corresponde à área sob a curva ROC (*Receiver Operating Characteristic*) utilizada para avaliar um classificador binário.

6.4 RESULTADOS

6.4.1 Análise de sensibilidade dos modelos

A primeira análise realizada no banco de dados foi avaliar o efeito da função de ativação escolhida (logística e ReLu) no desempenho das redes neurais. Assim, foram treinadas 49 redes neurais variando o número de neurônios na camada intermediária, de 2 a 50 neurônios. Os resultados mostram que a função de ativação ReLu forneceu alto valor de acurácia para os dados de treinamento e teste, com mínimo de 0,814 e máximo de 0,892 para treinamento, e mínimo de 0,783 e máximo de 0,878, para teste. Quando utilizada a função logística, estes valores caíram para mínimo de 0,810 e máximo de 0,833, para treinamento, e mínimo de 0,722 e máximo de 0,809, para teste (Figura 32). Conforme mencionado anteriormente, o valor máximo indicado na camada intermediária é de 15 de neurônios, e como mostra a Figura 32, aplicando a função ReLu, a estabilização acontece após 19 neurônios, o que ultrapassa o indicado por Netch-Nielsen (1987). Para a função logística, a estabilização acontece após 8 neurônios. Desta forma, a função de ativação logística, apesar de apresentar valores de acurácia mais baixos que

a função ReLu, estabiliza-se com um número de neurônios abaixo do máximo indicado, o que pode evidenciar que ela converge a resultados mais promissores a partir de redes de arquiteturas mais simples.

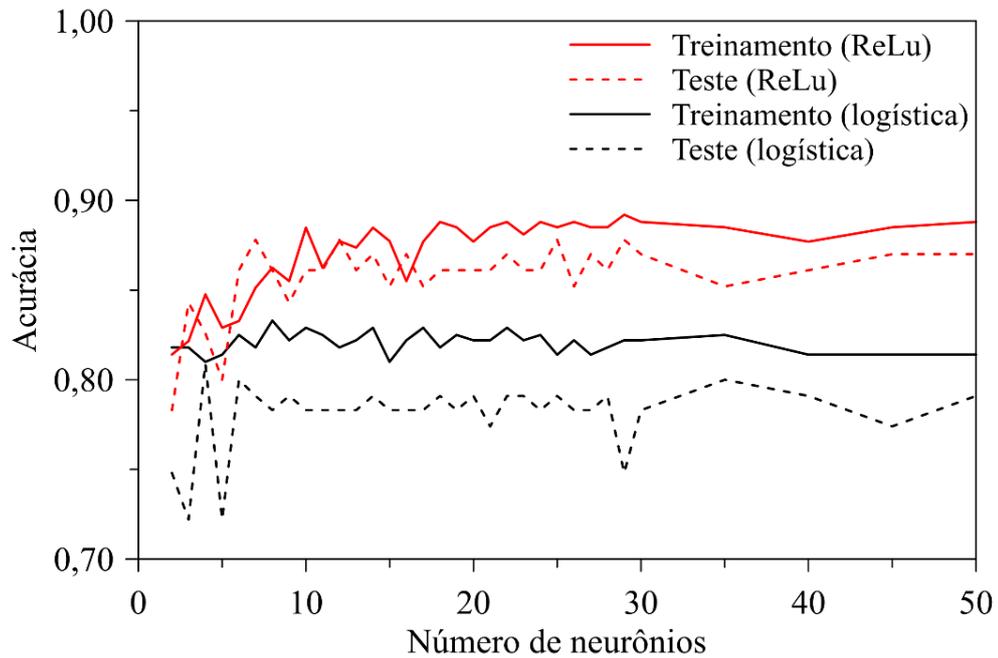


Figura 32: Variação da acurácia de treinamento e teste para diferentes números de neurônios na camada intermediária.

Após esta análise, foi implementado o método *random sampling*, com 30 reamostragens, para a rede neural com 8 neurônios na camada intermediária (7x8x2) e função de ativação logística. Novamente foi analisada a variação do desempenho das redes por meio dos valores de acurácia obtidos no treinamento e no teste. Observou-se que, apesar dos valores de acurácia serem satisfatórios, o modelo se apresenta instável, com valores oscilantes tanto para treinamento quanto para teste (Figura 33).

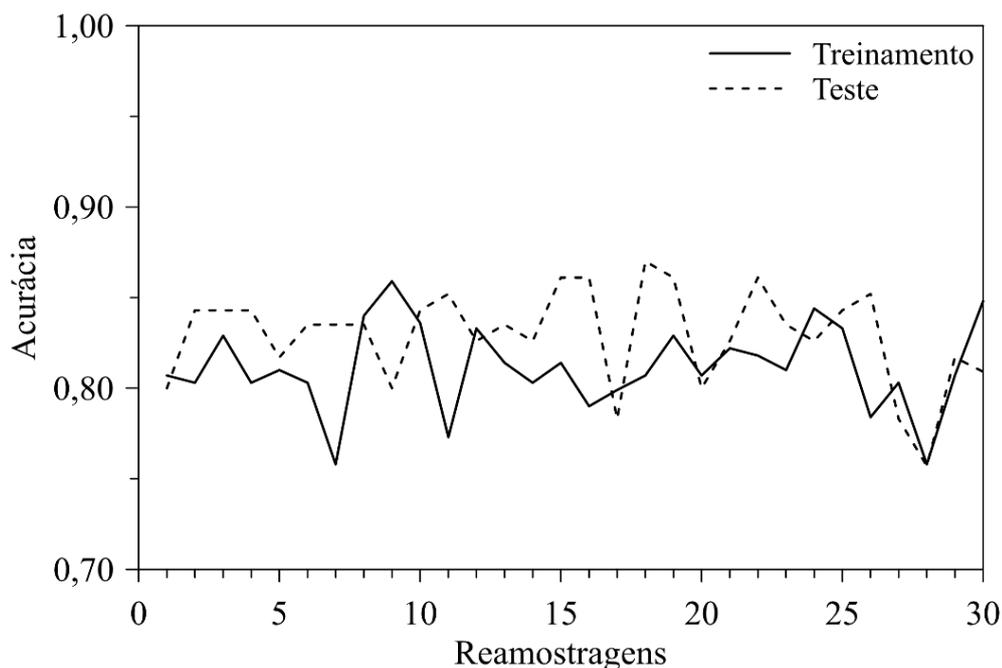


Figura 33: Variação da acurácia de treinamento e teste para 30 reamostragens aleatórias do banco de dados de treinamento.

Para o cálculo da suscetibilidade à ocorrência de cavidades ferríferas foi escolhida a rede com 8 neurônios na camada intermediária (7x8x2), que forneceu o menor valor de entropia cruzada (*cross-entropy*), de 0,108. A entropia cruzada pode ser usada como uma medida de erro quando o vetor de saída representa uma distribuição de probabilidade (Golik et al., 2013), e representa o quão afastado o valor obtido está do valor esperado. O desempenho do método preditivo foi avaliado por meio das métricas: acurácia, F1, precisão e recall, além dos valores de AUC correspondentes à taxa de sucesso e à taxa de precisão (Tabela 15).

Tabela 15: Métricas de avaliação de desempenho.

Métricas	AUC	Acurácia	Precisão	Recall	F1
Treinamento	0,886	0,833	0,815	0,869	0,841
Teste	0,866	0,783	0,795	0,783	0,782

É possível observar nos resultados da Tabela 15 que a métrica com maior queda entre o treinamento e o teste foi o Recall, com cerca de 9,9% . Isto indica que o modelo classifica mais

falsos negativos (FN) do que o esperado. Ou seja, pode haver uma tendência de classificar lugares onde existem cavidades como lugares onde não existem cavidades.

6.4.2 Suscetibilidade à ocorrência de cavidades ferríferas

A aplicação desta rede (7x8x2) resultou em um mapa de suscetibilidade a ocorrência de cavidades ferríferas (Figura 34). Utilizando-se de aprendizado supervisionado, a rede neural pode prever qual a possibilidade de pertencimento à uma das categorias: ter cavidade (target 1) e não ter cavidade (target 0).

Para determinar os intervalos das classes de suscetibilidade, o classificador Natural Break foi utilizado, o que resultou na representação de suscetibilidade a cavidades ferríferas por 4 classes: muito alta, alta, moderada e baixa (Figura 34).

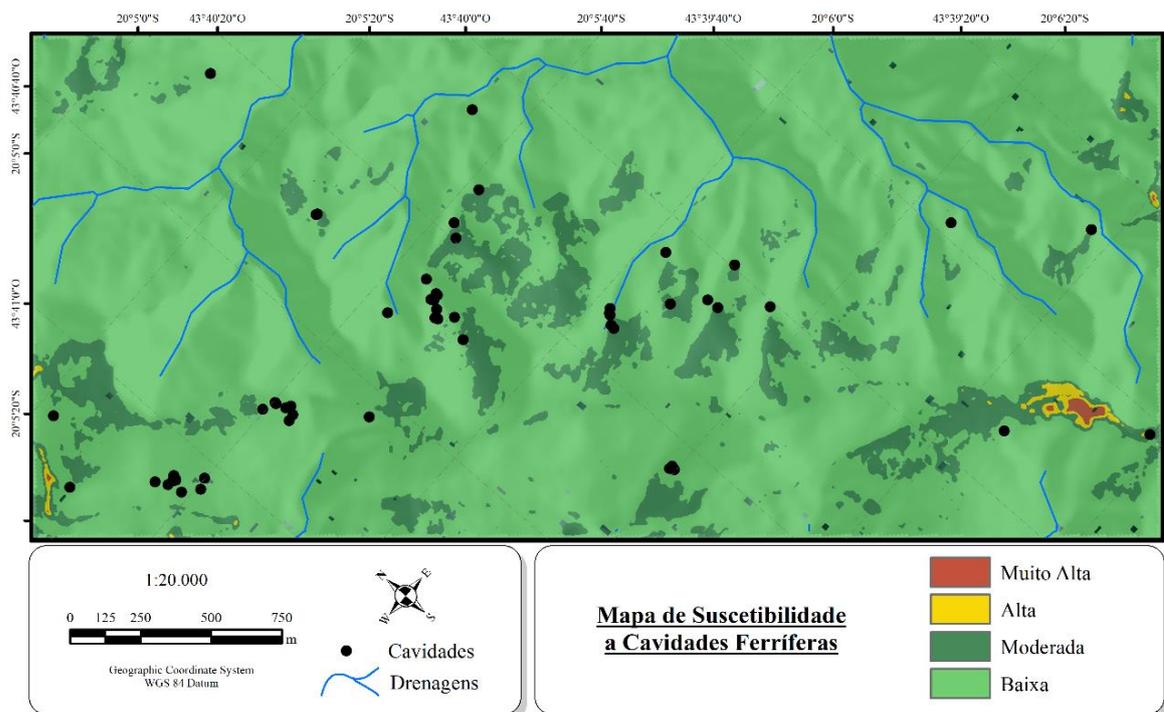


Figura 34: Mapa de suscetibilidade a ocorrência de cavidades ferríferas utilizando ANN.

Os resultados da predição de novos valores utilizando a rede treinada apresentou uma distribuição desbalanceada. Os valores mais altos de suscetibilidade foram priorizados na região onde existe solo exposto (vide Figura 29), sudeste do mapa, correspondente a classe muito alta. A classe alta ocupou uma pequena porção no entorno da classe muito alta. O restante

do mapa foi classificado majoritariamente como classe baixa, e dispersos nesta classe encontram-se alguns polígonos da classe moderada. Ao sobrepor o inventário de cavidades ao mapa de suscetibilidade, observou-se que estas se encontram, em sua maioria (96%), sobre a classe baixa.

O total de pixels do mapa de suscetibilidade à ocorrência de cavidades ferríferas é de 72004. Os valores de suscetibilidade correspondentes a cada pixel estão entre o intervalo de 0,01 a 0,28. Analisando-se a distribuição dos dados, observa-se que cerca de 60 mil pixels possuem valores abaixo de 0,02, o que corresponde a aproximadamente 83% do mapa. Isso demonstra que a variância dos dados é muito pequena, ou seja, há uma concentração muito alta de pixels em um intervalo muito pequeno de valores. Esse comportamento dificulta a utilização do classificador Natural Breaks para a separação de classes de suscetibilidade à cavidades ferríferas, pois ele se baseia na variabilidade dos dados buscando minimizar a variância dentro das classes e maximizar a variância entre as classes (Jenks, 1967). Por isso, este classificador não é recomendado para dados com baixa variância.

Portanto, o mapa, como elaborado e classificado, apresentou uma capacidade preditiva insatisfatória.

6.5 DISCUSSÃO DOS RESULTADOS

A utilização de Redes Neurais Artificiais se mostrou inadequada para o conjunto de dados desta pesquisa. O ajuste dos parâmetros foi realizado com base na concepção de uma rede neural que tinha por objetivo mapear a suscetibilidade a ocorrências de cavidades ferríferas. Isto significa que a escolha dos parâmetros foi fundamentada nas especificações de cada um deles, buscando encontrar um modelo com boa capacidade preditiva.

Para uma aplicação utilizando ANN, quanto mais complexa for a rede, em termos de tipos, arquiteturas e tipologias, mais dados de treinamento serão necessários para generalizar as soluções (Silva et al., 2010). Diante disso, buscou-se executar uma rede que tivesse um bom desempenho, mas que fosse simples, especialmente com o mínimo possível de neurônios na camada intermediária.

Do ponto de vista do desempenho durante o treinamento, a rede escolhida (7x8x2) obteve resultados satisfatórios apresentando uma alta taxa de sucesso (AUC = 0,886). Entretanto dois

pontos merecem destaques: (1) a oscilação dos valores de acurácia na análise por *random sampling*; (2) a queda de 9,9% do valor de recall entre o treinamento e o teste. A primeira observação demonstra que ao apresentar diferentes amostras de um mesmo conjunto de dados de treinamento, o modelo se mostrou instável. Em algumas reamostragens a acurácia do treinamento foi maior que a do teste, mas a maioria mostrou um comportamento inverso. Vale ressaltar que não se trata necessariamente de um *overfitting* porque a distância entre as curvas de treinamento e teste não foi muito significativa, menor que 0,1. Na segunda observação, o valor do recall para o teste diminuiu em relação ao treinamento. A depender do objetivo da aprendizagem, Powers (2020) explica que o recall pode ter um peso grande na avaliação de desempenho, principalmente quando o objetivo é identificar todos os casos de Verdadeiro Positivo (TP). Nesta situação, se a rede neural mostrar uma tendência de classificar Falsos Negativos (FN), o modelo classifica, em sua grande maioria, o *target* 1 (tem cavidade) como *target* 0 (não tem cavidade), prejudicando a predição da ocorrência de cavidades.

Ao mostrar novos dados à rede neural treinada, mesmo ela apresentando um bom desempenho, não houve generalização. Ou seja, a rede neural não possibilitou retirar todo o conhecimento necessário, a partir dos dados de treinamento, para classificar dados nunca vistos. Além disso, ela apresentou valores cuja variância é muito baixa, o que inviabilizou a utilização do classificador Natural Breaks.

Entende-se que a baixa capacidade de generalização pode ser consequência direta do tamanho do banco de dados. Alguns autores sugerem que um banco de dados de treinamento muito pequeno pode atrapalhar o processo de otimização, e que quanto maior o conjunto de dados de treinamento, melhor será a capacidade de generalização (Haykin, 2004; Chi et al., 2008; Silva et al., 2010; Feng et al., 2019). De fato, o banco de dados da pesquisa foi pequeno (384 instâncias) e, mesmo utilizando a técnica de validação cruzada, indicada para esse tipo de circunstância, a rede neural demonstrou ter dificuldade de gerar soluções para dados não vistos.

Por fim, quando comparados estes resultados com aqueles apresentados por Nola e Bacellar (2021, 2022), conclui-se que a técnica RF possui maior facilidade de aplicação e interpretação, como apontado por Bergen et al. (2019). A técnica AHP possibilitou menor gasto computacional, entretanto, a sua aplicação foi dificultada pela necessidade do emprego de lógica fuzzy para padronização dos fatores. A ANN mostrou-se a técnica mais complexa e com maior dificuldade de interpretação.

As técnicas RF e ANN apresentaram valores satisfatórios, com destaque para a primeira, que forneceu os valores mais altos para as taxas de sucesso e de predição, respectivamente 0,954 e 0,957. Entretanto, apesar do desempenho da aprendizagem por ANN ser satisfatório, sua capacidade de generalização foi insatisfatória.

A partir das características do banco de dados utilizado nesta pesquisa, os métodos para aplicação das técnicas AHP e RF demonstraram ser facilmente replicáveis. Todavia, na aplicação de AHP, se faz necessária a consulta de um grupo de especialistas que julgam a importância de cada fator, culminando em uma análise subjetiva dos pesos.

6.6 CONCLUSÃO

Nesta pesquisa buscou-se avaliar o emprego de ANN no mapeamento da suscetibilidade à ocorrência de cavidades ferríferas. Os resultados indicam que a rede treinada não foi capaz de identificar as relações entre os fatores condicionantes, apesar de apresentar durante o treinamento uma boa capacidade preditiva.

Para melhorar a generalização de ANN, recomenda-se aumentar o banco de dados de treinamento. Entretanto, a prospecção espeleológica para identificar coletar e montar um grande conjunto de dados suficientes para tal aplicação é um desafio. Uma alternativa é buscar novas ferramentas capazes de lidar com conjunto de dados pequenos, como a utilização de redes neurais profundas com pré-treinamento (Feng et al., 2019). Todavia, utilizar redes neurais profundas é ainda mais complexo e pode gerar resultados inferiores aos obtidos por redes neurais rasas (Malinov et al., 2001). Outra sugestão é diminuir o número de variáveis para melhorar o modelo, porque isso diminui a complexidade da rede (Silva et al., 2010).

Os resultados aqui apresentados não foram consistentes e concordantes com os demonstrados pelas técnicas AHP e RF (Nola e Bacellar, 2021; 2022). A ANN apresentou maior complexidade, custo computacional e dificuldade de interpretação. Mesmo sendo uma técnica robusta e amplamente utilizada, o uso de um limitado conjunto de informações rotuladas pode ter contribuído para a baixa eficiência desta técnica.

CAPÍTULO 7

CONCLUSÃO

Esta pesquisa apresentou o estudo da suscetibilidade a ocorrência de cavidades ferríferas por meio de técnicas de análise multicritério e aprendizado de máquina, que foram avaliadas quanto a sua aplicabilidade e capacidade preditiva. Para realizar essa pesquisa foi selecionada uma área de aproximadamente 7,0 km² na Serra do Gandarela, pertencente ao Quadrilátero Ferrífero/MG, onde existe uma alta concentração de cavidades já mapeadas. O banco de dados utilizado foi composto por 7 fatores condicionantes a ocorrência de cavidades ferríferas. Esses fatores foram inicialmente levantados tendo em conta algumas características geológicas, morfológicas e hidrológicas das cavidades que podem estar relacionadas à gênese e ao desenvolvimento de cavidades em materiais ricos em ferro. Além disso, foram coletadas em um banco de dados público as localizações das cavidades ferríferas na área de estudo.

A utilização de AHP, com padronização de variáveis por lógica fuzzy se mostrou adequada para o mapeamento da suscetibilidade à ocorrência de cavidades ferríferas. Esta técnica se mostrou de fácil interpretação, além de ter baixo gasto computacional. Por outro lado, o tempo gasto para estruturação hierárquica dos fatores foi excessivo e, o julgamento para definição dos pesos pelos especialistas, apesar de ser baseado em experiências acumuladas, inseriu certo grau de incerteza nos resultados devido a sua subjetividade. Estas incertezas podem acontecer principalmente em áreas do conhecimento onde o conceito que rege o problema ainda não está bem esclarecido entre os especialistas.

A técnica de RF apresentou maior facilidade de aplicação e interpretação, quando comparada as outras técnicas avaliadas nessa pesquisa. Os resultados obtidos revelaram elevadas taxas de sucesso e de predição. O mapa de suscetibilidade à ocorrência de cavidades ferríferas produzido por meio de RF, apresentou alta capacidade preditiva, o que pode ser observado pelo alto valor de *Frequency Ratio*, para as classes alta e muito alta. Além disso, foi possível constatar por meio do *Gini importance* que o NDVI e o HAND são os fatores que mais contribuíram no processo de classificação, o que ressalta a relação existente entre esses fatores e o desenvolvimento de cavidades ferríferas.

A técnica de ANN foi a técnica mais complexa utilizada nesta pesquisa, apresentando maior gasto computacional e maior dificuldade de interpretação dos resultados. A sua baixa capacidade de identificar as relações entre os fatores condicionantes e seus rótulos, resultou em um modelo com capacidade de generalização insatisfatória, mesmo tendo apresentado uma boa capacidade preditiva durante o treinamento. Assim, a rede neural avaliada foi ineficaz ao classificar dados nunca vistos, o que possivelmente pode estar relacionado ao limitado conjunto de dados para treinamento.

Os resultados desta pesquisa mostraram que AHP e RF são técnicas capazes de realizar uma previsão rápida, remota e de baixo custo de áreas suscetíveis à ocorrência de cavidades ferríferas. Dessa forma, o trabalho apresentado pode auxiliar os tomadores de decisão, planejadores e autoridades competentes na identificação de locais alvo para a prospecção espeleológica, podendo ser aplicada em outras áreas com características geológicas semelhantes.

A presente pesquisa foi pioneira na avaliação de técnicas preditivas no mapeamento da suscetibilidade a ocorrência de cavidades ferríferas. Diante desse cenário foram observadas algumas questões que necessitam de trabalhos futuros, visando promover avanços científicos nessa área do conhecimento, portanto, recomenda-se:

- A agregação de outros fatores, tais como dados aerogeofísicos ou *knickpoints*, com o objetivo de avaliar novas relações entre esses fatores e a ocorrência de cavidades ferríferas;
- A expansão do banco de dados de treinamento, a fim de melhorar o mapeamento das áreas de alta e muito alta suscetibilidade;
- Avaliar outras formas de amostragem dos pontos onde não existem cavidades, para controlar e refinar os valores mais efetivos na classificação do target 0 (não cavidade);
- Avaliar o efeito da remoção dos fatores de menor importância, reduzindo a complexidade do banco de dados, visando observar alterações no desempenho das técnicas preditivas.

REFERÊNCIAS BIBLIOGRÁFICAS

- Adition, A., Kubota, T., & Shinohara, Y., 2018. Comparison of GIS-based landslide susceptibility models using frequency ratio, logistic regression, and artificial neural network in a tertiary region of Ambon, Indonesia. *Geomorphology*, 318, 101-111.
- Agapov, I.A., Khlebalin, I.Y., Lyakhnitsky, Y.S. 2016. Caves of the Sokanlinna granite massif, Vyborg District (Leningrad Region, Russia). *Pseudokarst Commission Newsletter*, 26: 15-22.
- Ahmad, M., Tang, X. and Ahmad, F. 2020. Evaluation of Liquefaction-Induced Settlement Using Random Forest and REP Tree Models: Taking Pohang Earthquake as a Case of Illustration. In *Natural Hazards-Impacts, Adjustments & Resilience*. IntechOpen. doi: 10.5772/intechopen.94274
- Albuquerque, A.R., Angélica, R.S., Gonçalves, D.F., Paz, S.P. 2018. Phosphate speleothems in caves developed in iron ores and laterites of the Carajás Mineral Province (Brazil) and a new occurrence of spheniscidite. *International Journal of Speleology*, 47(1): 53-67.
- Alkmim, F.F., Marshak, S. 1998. Transamazonian Orogeny in the Southern São Francisco Craton Region, Minas Gerais, Brazil: evidence for Paleoproterozoic collision and collapse in the Quadrilátero Ferrífero. *Precambrian Research*, 90: 29-58.
- Arabameri, A., Pourghasemi, H.R. 2019. Spatial modeling of gully erosion using linear and quadratic discriminant analyses in GIS and R. In: Pourghasemi, H.R., Gokceoglu, C. (Eds.), *Spatial Modeling in GIS and R for Earth and Environmental Sciences*. Elsevier, pp. 299-321.
- Arthi, J., & Akoramurthy, B. 2018. Extrapolation and Visualization of NPA Using Feature Based Random Forest Algorithm in Indian Banks. In *Progress in Computing, Analytics and Networking*. Springer, Singapore, pp. 601-612.
- Aubrecht, R., Láncoz, T., Šmída, B., Brewer-Carías, C., Mayoral, F., Schlögl, J., ... & Gregor, M. 2008. Venezuelan sandstone caves: a new view on their genesis, hydrogeology and speleothems. *Geologia Croatica*, 61(2-3): 345-362.
- Auler A. S., Piló, L. B. 2005. Introdução às cavidades em minério de ferro e canga. *O Carste*, 17(3): 70-72.
- Auler, A., Piló L., Parker, C., Senko, J., Sasowsky, I., Barton, H. 2014. Hypogene cave patterns in iron ore caves: Convergence of forms or processes? In: Klimchouk A.B., Sasowsky I.D., Mylroie J., Engel S.A., Engel A.S. (Eds.). *Hypogene Cave Morphologies*, Karst Water Institute, Virginia, pp. 15-19.
- Auler, A.S., Farrant, A.R. 1996. A brief introduction to karst and caves in Brazil. *Proceedings of the University of Bristol Speleological Society*, 20(3): 187-200.
- Auler, A.S., Parker, C.W., Barton, H.A., Soares, G.A. 2019. Iron formation caves: Genesis and ecology. In: White, W.B., Culver, D. C., Pipan, T. (Eds.), *Encyclopedia of Caves*. Academic Press, pp. 559-566.
- Auler, A.S., Piló, L.B. 2015. Caves and mining in Brazil: the dilemma of cave preservation within a mining context. In: Andreo, B., Carrasco, F., Durán, J., Jiménez, P., LaMoreaux, J. (Eds.), *Hydrogeological and Environmental Investigations in Karst Systems*. Springer, Heidelberg, pp. 487-496.

- Bacellar, L.A.P., Coelho Netto, A.L., Lacerda, W.A. 2005. Controlling factors of gullying in the Maracujá Catchment, southeastern Brazil. *Earth Surface Processes and Landforms*, 30(11): 1369-1385. doi:10.1002/esp.1193.
- Baeta, A., Piló, H. 2005. Arqueologia nas cavernas e suportes ferruginosos. (*portuguese*) In: Ruchkys, O. A., Rasteiro, M., Travassos, L.P., Faria, L. (Eds.), *Patrimônio Espeleológico em Rochas Ferruginosas: Propostas Para Sua Conservação no Quadrilátero Ferrífero, Minas Gerais*. Sociedade Brasileira de Espeleologia, Campinas, pp. 210-239.
- Band, S. S., Janizadeh, S., Chandra Pal, S., Saha, A., Chakraborty, R., Shokri, M., & Mosavi, A., 2020. Novel ensemble approach of deep learning neural network (DLNN) model and particle swarm optimization (PSO) algorithm for prediction of gully erosion susceptibility. *Sensors*, 20(19), 5609.
- Barbosa, M.R., da Silva, A.D., De Paula, R.G., Dutra, G.M., Barata, A., Brandi, I.V., da Silva, C.R., Osborne, R.A.L. 2019. Breakdown mechanisms in iron caves. An example from Brazil. *International Journal of Speleology*, 48(2): 179-190.
- Beisiegel, V. R. et al. 1973. Geologia e recursos minerais da Serra dos Carajás. *Revista Brasileira de Geociências*, 3(4): 215-242.
- Beucher, A. et al., 2013. Artificial neural network for acid sulfate soil mapping: application to the Sirppujoki River catchment area, south-western Finland. *Journal of Geochemical Exploration*, 125: 46-55.
- Bittencourt, J.S., Vasconcelos, A.G., Carmo, F.F., Buchmann, F.S. 2015. Registro paleontológico em caverna desenvolvida em formações ferríferas na Serra do Gandarela. (*portuguese*) In: Ruchkys, O.A., Rasteiro, M., Travassos, L.P., Faria, L. (Eds.), *Patrimônio Espeleológico em Rochas Ferruginosas: Propostas Para Sua Conservação no Quadrilátero Ferrífero, Minas Gerais*. Sociedade Brasileira de Espeleologia, Campinas, pp. 191-209.
- Bo, L., Yi-Fan, Z., Bei-Bei, Z., & Xian-Qing, W., 2018. A risk evaluation model for karst groundwater pollution based on geographic information system and artificial neural network applications. *Environmental Earth Sciences*, 77(9), 1-14.
- Boulesteix, A. L., Janitza, S., Kruppa, J., & König, I. R. 2012. Overview of random forest methodology and practical guidance with emphasis on computational biology and bioinformatics. *Wiley Interdisciplinary Reviews: Data Mining and Knowledge Discovery*, 2(6): 493-507.
- Bragagnolo, L., da Silva, R. V., & Grzybowski, J. M. V., 2020. Artificial neural network ensembles applied to the mapping of landslide susceptibility. *Catena*, 184, 104240.
- Brandi, I.V., Barbosa, M.R., Calux, A.S., Araujo, R.N. 2015. Geotechnical knowledge applied to the predictable impacts of physical integrity of natural caves, Carajás - PA (*portuguese*). In: Rasteiro, M.A., Sallun Filho, W. (Eds.), *Proceedings of the 33th Brazilian Congress of Speleology*, Brazilian Speleological Society, Eldorado, pp. 533-541.
- Brazilian Government. Decreto de 13 de outubro de 2014 (*Brazilian law*). Cria o Parque Nacional da Serra do Gandarela, localizado nos Municípios de Nova Lima, Raposos, Caeté, Santa Bárbara, Mariana, Ouro Preto, Itabirito e Rio Acima, Estado de Minas Gerais. Brasília, DF, out. 2014.
- Breiman L. 2001. Random forests. *Machine Learning*, 45: 5-32.

- Breiman, L., Friedman J. H., Olshen R. A., Stone C. J. 1984. Classification and Regression Trees. CRC Press, 1984.
- Breiman. L. 1996. Bagging predictors. *Machine Learning*, 24: 123-140.
- Brito, M.M., Weber, E.J., Silva Filho, L.C.P. 2017. Multi-criteria analysis applied to landslide susceptibility mapping. *Revista Brasileira de Geomorfologia*, 18(4): 719-735.
- Cadamuro, A. L. M. 2001. Relatório Final da Fase I: Estudos Geológicos e Hidrogeológicos relativos à Gruta dos Ecos, Cocalzinho – GO. Projeto BRA/00/009, IBAMA, Produto 4, Brasília. 18p.
- Calligaris, C., Devoto, S., Galve, J., Zini, L., Pérez Peña, J. 2017. Integration of multi-criteria and nearest neighbour analysis with kernel density functions for improving sinkhole susceptibility models: the case study of Enemonzo (NE Italy). *International Journal of Speleology*, 46(2): 191-204.
- Calligaris, C., Poretti, G., Tariq, S., Melis, M.T. 2013. First steps towards a landslide inventory map of the Central Karakoram National Park. *European Journal of Remote Sensing*, 46(1): 272-287. doi: 10.5721/EuJRS20134615
- Calux, A.S. 2013. Gênese e desenvolvimento de cavidades naturais subterrâneas em Formação Ferrífera no Quadrilátero Ferrífero, Minas Gerais. Dissertação de mestrado, Programa de pós-graduação em Geografia. Universidade Federal de Minas Gerais, Belo Horizonte, MG.
- Calux, A.S., Cassimiro, R. 2015. Geoespeleologia das cavernas em rochas ferríferas: Aspectos dimensionais, morfológicos, hidrológicos e sedimentares. (*portuguese*) In: Ruchkys, O.A., Rasteiro, M., Travassos, L.P., Faria, L. (Eds.), Patrimônio Espeleológico em Rochas Ferruginosas: Propostas Para Sua Conservação no Quadrilátero Ferrífero, Minas Gerais. Sociedade Brasileira de Espeleologia, Campinas, pp. 132-157.
- Calux, A.S., Cassimiro, R., Salgado, A. 2019. Caves in iron formations in the Quadrilátero Ferrífero, Minas Gerais, southeastern Brazil: lithological, morphological and hydrological settings and speleogenesis. *Zeitschrift fur Geomorphologie*, 62(2): 125-144.
- Carmo, F. F.; Carmo, F. F.; Salgado, A. A. R.; Jacobi, C. M. 2011. Novo sítio espeleológico em sistemas ferruginosos, no Vale do Rio Peixe Bravo, norte de Minas Gerais, Brasil. *Espeleo-Tema*, 22(1): 25-39.
- Carmo, F.F., Kamino, L.H.Y. 2017. O Vale do Rio Peixe Bravo: Ilhas de Ferro no Sertão Mineiro. 3i Editora, Belo Horizonte, 208 p.
- Casas, A.M., Cortes, A.L., Maestro, A., Soriano, M.A., Riaguas, A., Bernal, J. 2000. Lindens: a program for lineament length and density analysis. *Computers & Geosciences*, 26(10): 1011-1022.
- CECAV. National Cave Research and Conservation Center. Brazilian Speleological Information Registry. <https://www.icmbio.gov.br/cecav/canie.html> [accessed: February 10, 2019].
- Çelik, S.; Tan, Ö. 2005. Determination of preconsolidation pressure with artificial neural network. *Civil Engineering and Environmental Systems*, 22(4): 217-231.
- Chakraborty, S., Mukhopadhyay, S. 2019. Assessing flood risk using analytical hierarchy process (AHP) and geographical information system (GIS): application in Coochbehar district of West Bengal, India. *Natural Hazards*, 99(1): 247-274.

- Cho, S. E., 2009. Probabilistic stability analyses of slopes using the ANN-based response surface. *Computers and Geotechnics*, 36(5), 787-797
- Ciampalini, A., Garfagnoli, F., Antonielli, B., Moretti, S., Righini, G. 2013. Remote sensing techniques using Landsat ETM+ applied to the detection of iron ore deposits in Western Africa. *Arabian Journal of Geosciences*, 6(11): 4529-4546.
- Clark, C.D., Wilson, C. 1994. Spatial analysis of lineaments. *Computers and Geosciences*, 20(8): 1237-1258.
- CNC - Cadastro Nacional de Cavernas do Brasil. 2021. Disponível em: <http://www.cavernas.org.br/cnc/>. [acessado em 10 de junho 2021].
- Coelho, A.; Piló L.B.; Auler, A. & Bessi, R. 2010. Espeleologia da área do Projeto Apolo, Quadrilátero Ferrífero, MG. *Carste Consultores Associados*, 179 p.
- Das, B., Pal, S.C. 2019. Combination of GIS and fuzzy-AHP for delineating groundwater recharge potential zones in the critical Goghat-II block of West Bengal, India. *HydroResearch*, 2: 21-30.
- Demsar J., Curk T., Erjavec A., Gorup C., Hocevar T., Milutinovic M., Mozina M., Polajnar M., Toplak M., Staric A., Stajdohar M., Umek L., Zagar L., Zbontar J., Zitnik M., Zupan B. 2013. Orange: Data Mining Toolbox in Python. *Journal of Machine Learning Research*, 14: 2349–2353.
- Di Napoli, M., Carotenuto, F., Cevasco, A., Confuorto, P., Di Martire, D., Firpo, M., Pepe, G., Raso, E., Calcaterra, D. 2020. Machine learning ensemble modelling as a tool to improve landslide susceptibility mapping reliability. *Landslides*, 17(8): 1897-1914.
- Dixey, F. 1920. Notes on laterization in Sierra Leone. *Geol. Mag.*, 57, 671: 211-220.
- Dodgson, J.S., Spackman, M., Pearman, A., Phillips, L.D. 2009. Multi-criteria analysis: a manual. Department for Communities and Local Government, London, 168p.
- Dorr, J.V.N. 1969. Physiographic stratigraphic and structural development of the Quadrilátero Ferrífero, Minas Gerais, Brazil. *DNPM/USGS*, Washington, 109 p.
- Ducart, D.F., Silva, A.M., Toledo, C.L.B., Assis, L.M.D. 2016. Mapping iron oxides with Landsat-8/OLI and EO-1/Hyperion imagery from the Serra Norte iron deposits in the Carajás Mineral Province, Brazil. *Brazilian Journal of Geology*, 46(3): 331-349.
- Dutra, G.M. 2013. Synthesis of cavities genesis in iron formation and correlates lithologies (*portuguese*) In: Rasteiro, M.A., Morato, L. (Eds.), *Proceedings of the 32th Brazilian Congress of Speleology*, Brazilian Speleology Society, Barreiras, pp. 415-426.
- Dutra, G.M. 2015. Gênese e Desenvolvimento de Cavidades Naturais Subterrâneas em Formações Ferríferas. (*portuguese*) In: Ruchkys, O.A., Rasteiro, M., Travassos, L.P., Faria, L. (Eds.), *Patrimônio Espeleológico em Rochas Ferruginosas: Propostas Para Sua Conservação no Quadrilátero Ferrífero, Minas Gerais*. Sociedade Brasileira de Espeleologia, Campinas, pp. 158-173.
- Dutra, G.M. 2017. Análise de susceptibilidade de duas cavidades em litologia de ferro na Serra do Gandarela, MG. Estudo de caso: AP_0009 e AP_0038. Dissertação de mestrado, Programa de pós-graduação em Geotecnia. NUGEO. Universidade Federal de Ouro Preto, Ouro Preto, MG.

- Eastman, J.R. 2012. IDRISI Selva GIS and image processing software version 17.0. Clark Labs, Massachusetts. (<https://clarklabs.org/download/>)
- Elmahdy, S.I., Mohamed, M.M., Ali, T.A., Abdalla, J.E.D. and Abouleish, M. 2020. Land subsidence and sinkholes susceptibility mapping and analysis using random forest and frequency ratio models in Al Ain, UAE. *Geocarto International*, pp. 1-17.
- ENVI, 2001. ENVI Programmer's Guide. Research System, Inc. (<https://lost-contact.mit.edu/afs/rsi/envi/docs/proguid>).
- ESA, 2015. Sentinel-2 User Handbook. ESA (European Space Agency) Standard Document. (https://sentinel.esa.int/documents/247904/685211/Sentinel-2_User_Handbook)
- ESRI, 2012. ArcGIS 10.1. Environmental Systems Research Institute, Redlands, CA, USA.
- Falah, F., Rahmati, O., Rostami, M., Ahmadisharaf, E., Daliakopoulos, I. N., & Pourghasemi, H. R. (2019). Artificial neural networks for flood susceptibility mapping in data-scarce urban areas. In *Spatial modeling in GIS and R for Earth and Environmental Sciences* (pp. 323-336). Elsevier.
- Fayyad, U.; Piatetsky-Shapiro, G.; Smyth, P. 1996. The KDD process for extracting useful knowledge from volumes of data. *Communications of the ACM*, 39(11): 27-34.
- Feng, S., Zhou, H., & Dong, H., 2019. Using deep neural network with small dataset to predict material defects. *Materials & Design*, 162, 300-310.
- Fernandes, G.W. 2016. The shady future of the rupestrian grassland: major threats to conservation and challenges in the Anthropocene. In: Fernandes, G. W. (Ed.), *Ecology and conservation of mountaintop grasslands in Brazil*. Springer, Switzerland, pp. 545-561.
- Ferreira, R.L. 2005. A vida subterrânea nos campos ferruginosos. *O Carste*, 3(17): 106-115.
- Figueiredo L. A. V. 2011. História da espeleologia brasileira: protagonismo e atualização cronológica. *anais do 31 Congresso Brasileiro de Espeleologia Ponta Grossa-PR, 21-24 de julho de 2011 – Sociedade Brasileira de Espeleologia*. pp. 379-395.
- Florinsky, I.V. 2016. *Digital terrain analysis in soil science and geology*. Academic Press, London, 486 p.
- Florinsky, I.V., Kuryakova, G.A. 1996. Influence of topography on some vegetation cover properties. *Catena*, 27: 123-141.
- Frank, H.T., Buchmann, F.S.C., Lima, L.G., Fornari, M., Caron, F., Lopes, R.P. 2012. Cenozoic Vertebrate Tunnels in Southern Brazil. *Ichnology of Latin America*, 2: 141-158.
- Freeze, R.A., Cherry, J.A. 1979. *Groundwater*. Englewood Cliffs, Prentice-Hall, New Jersey, 604 p.
- Ganem, R. S. 2009. As cavidades naturais subterrâneas e o Decreto nº 6.640/2008. *Biblioteca Digital da Câmara dos Deputados, Brasília/DF*, 34 p.
- Gautier, E.F. 1906. *Madagascar Essai de géographie physique*. A. Challamel, Paris, 428 p.
- Gharari, S., Hrachowitz, M., Fenicia, F., Savenije, H.H.G. 2011. Hydrological landscape classification: investigating the performance of HAND based landscape classifications in a central European meso-scale catchment. *Hydrology and Earth System Sciences*, 15: 3275-3291.

- Ghazifard, A., Nikoobakht, S., & Azarafza, M., 2016. Municipal Waste Landfill Site Selection based on Environmental, Geological and Geotechnical Multi-criteria: A Case Study. *Iranian Journal of Environmental Technology*, 2(1), 49-67.
- Ghosh, A., Kar, S.K. 2018. Application of analytical hierarchy process (AHP) for flood risk assessment: a case study in Malda district of West Bengal, India. *Natural Hazards*, 94: 349-368.
- Ghosh, J.K., Bhattacharya, D.N., Sharma, S.K. 2012. Fuzzy knowledge based GIS for zonation of landslide susceptibility. In: Banerjee S (ed.), *Applications of chaos and nonlinear dynamics in science and engineering*. Springer, Heidelberg, pp. 21-37.
- Golik, P., Doetsch, P., & Ney, H. (2013, August). Cross-entropy vs. squared error training: a theoretical and experimental comparison. In *Interspeech* (Vol. 13, pp. 1756-1760).
- Gong, F. Q., Li, X. B., & Deng, J. (2006). Probabilistic distribution of geotechnical parameters by using AHP prior distribution fusion method. *Yantu Gongcheng Xuebao*(Chinese Journal of Geotechnical Engineering), 28(10), 1313-1318.
- Gopinathan, P., Parthiban, S., Magendran, T., Al-Quraishi, A.M.F., Singh, A.K., Singh, P.K. 2020. Mapping of ferric (Fe^{3+}) and ferrous (Fe^{2+}) iron oxides distribution using band ratio techniques with ASTER data and geochemistry of Kanjamalai and Godumalai, Tamil Nadu, south India. *Remote Sensing Applications: Society and Environment*, 18, 100306.
- Grimes, K., Spate, A. 2008. Laterite karst. *ACKM Journal*, 73: 49-52.
- Guild, P.W. 1957. *Geology and mineral resources of the Congonhas district, Minas Gerais, Brazil*. Washington, USGS/DNPM, Professional Paper 290, 89 p.
- Han, J.; Kamber, M.; Pei, J. 2011. *Data mining: concepts and techniques*. Elsevier, 703 p.
- Hasekioğullari, G.D., Ercanoglu, M. 2012. A new approach to use AHP in landslide susceptibility mapping: a case study at Yenice (Karabuk, NW Turkey). *Natural Hazards*, 63(2): 1157-1179.
- Haykin, S. 2004. *Neural Network: A comprehensive foundation*. Neural networks, v. 2, 41 p.
- Hecht-Nielsen, R. (1987, June). Kolmogorov's mapping neural network existence theorem. In *Proceedings of the international conference on Neural Networks* (Vol. 3, pp. 11-14). IEEE Press New York.
- Holben, B.N. 1986. Characteristics of maximum value composite images from temporal AVHRR data. *International Journal of Remote Sensing*, 7(11): 1417-1434.
- Holler, C. 2019. Pseudokarst. In: White, W.B., Culver, D.C., Pipan, T. (Eds.), *Encyclopedia of Caves*. Academic Press, pp. 836-849.
- Huggett, R.J. 2011. *Fundamentals of Geomorphology*. Routledge Taylor and Francis Group, London and New York, 3rd edition, 516 p.
- Hwang, C.L., Yoon, K.P. 1981. *Multiple attribute decision making: Methods and applications*. Springer-Verlag, Berlin/Heidelberg/New York.
- Iniesta, L. F. M., & Ferreira, R. L. 2013. The first troglobitic Pseudonannolene from Brazilian iron ore caves (Spirostreptida: Pseudonannolenidae). *Zootaxa*, 3669(1): 85-95.
- INPE, 2019. TerraView Software Copyright © 2001-2020. Brazilian National Institute for Space Research. (<https://dpi.inpe.br/terraview>).

- Ishizaka, A.; Nemery, P. 2013. Multi-criteria decision analysis: methods and software. John Wiley & Sons.
- Jamshidi, J., Ataei, M., Sereshki, F., & Seyed, M. J., 2009. The application of AHP approach to selection of optimum underground mining method, case study: Jajarm Bauxite Mine (Iran). *Archives of Mining Sciences*, 54(1), 103-117.
- Janitza, S., & Hornung, R. 2018. On the overestimation of random forest's out-of-bag error. *PloS one*, 13(8), e0201904.
- Jenks G F, 1967. The Data Model Concept in Statistical Mapping. *International Yearbook of Cartography*, 7, 186-190.
- Jensen, J.R. 2009. Remote sensing of the environment: An earth resource perspective. Pearson Education India, 592 p.
- Joshi, R., Gupte, R., & Saravanan, P. 2018. A random forest approach for predicting online buying behavior of Indian customers. *Theoretical Economics Letters*, 8(03): 448.
- Jutson, J.T. 1914. Geological sketch map of Kurnalpi, N.E. Coolgardie. *Geol. Surv. Of Western Australia. Bulletin* 59.
- Kadam, A., Karnewar, A.S., Umrikar, B., Sankhua, R.N. 2019. Hydrological response-based watershed prioritization in semiarid, basaltic region of western India using frequency ratio, fuzzy logic and AHP method. *Environment, Development and Sustainability*, 21(4): 1809-1833.
- Kalantar, B., Ueda, N., Saeidi, V., Janizadeh, S., Shabani, F., Ahmadi, K., & Shabani, F., 2021. Deep neural network utilizing remote sensing datasets for flood hazard susceptibility mapping in Brisbane, Australia. *Remote Sensing*, 13(13), 2638.
- Kalogirou, S. A. 2000. Applications of artificial neural-networks for energy systems. *Applied Energy*, 67: 17-35.
- Kamp, U., Growley, B.J., Khattak, G.A., Owen, L.A. 2008. GIS-based landslide susceptibility mapping for the 2005 Kashmir earthquake region. *Geomorphology*, 101: 631-642.
- Karmann, I., Sánchez, L. E., & Fairchild, T. R. 2001. Caverna dos Ecos (Central Brazil): Genesis and geomorphologic context of a cave developed in schist, quartzite, and marble. *Journal of Cave and Karst Studies*, 63(1): 41-47.
- Kayastha, P., Dhital, M.R., De Smedt, F. 2013. Application of the analytical hierarchy process (AHP) for landslide susceptibility mapping: a case study from the Tinau watershed, west Nepal. *Computers & Geosciences*, 52: 398-408.
- Kiran, S. and Lal, B. 2016. Modelling of soil shear strength using neural network approach. *Electronic Journal of Geotechnical Engineering*, 21: 3751-3771.
- Koch, W., Meng, Y., Shah, M., Chang, W., & Yu, X., 2013. Predicting nerve guidance conduit performance for peripheral nerve regeneration using bootstrap aggregated neural networks. In *The 2013 International Joint Conference on Neural Networks (IJCNN)* (pp. 1-7). IEEE.
- Kohestani, V.R., Hassanlourad, M. & Ardakani, A. 2015. Evaluation of liquefaction potential based on CPT data using random forest. *Natural Hazards* 79: 1079-1089.
- Kornejady, A., Ownegh, M., Bahremand, A. 2017. Landslide susceptibility assessment using maximum entropy model with two different data sampling methods. *Catena*, 152: 144-162.

- Kovačević, M., Ivanišević, N., Dašić, T., & Marković, L., 2018. Application of artificial neural networks for hydrological modelling in karst. *Građevinar*, 70(01.), 1-10.
- Kumar, R., Anbalagan, R. 2016. Landslide susceptibility mapping using analytical hierarchy process (AHP) in Tehri reservoir rim region, Uttarakhand. *Journal of the Geological Society of India*, 87(3): 271-286.
- Kurup, P.U.; Dudani, N.K. 2002. Neural networks for profiling stress history of clays from PCPT data. *Journal of Geotechnical and Geoenvironmental Engineering*, 128 (7): 569-579.
- Lamounier, W.L., Carvalho, V.L.M., Salgado, A.A.R., Marent, B.R. 2010. A influência da litologia na distribuição da cobertura vegetal e no uso do solo na Serra do Gandarela, Quadrilátero Ferrífero - MG. *Geografias* 6(1): 152-165.
- Laureano, F.V., 1998. O registro sedimentar clástico associado aos sistemas de cavernas Lapa Doce e Torrinha, Município de Iraquara, Chapada Diamantina (BA). Tese de doutorado, Programa de pós-graduação em geoquímica e geotectônica. Universidade de São Paulo, São Paulo, SP.
- Lee, M.J., Kang, J.E., Jeon, S. 2012. Application of frequency ratio model and validation for predictive flooded area susceptibility mapping using GIS. In: *Proceedings of the 2012 IEEE International Geoscience and Remote Sensing Symposium*. IEEE, pp. 895-898.
- Lee, S., Pradhan, B. 2007. Landslide hazard mapping at Selangor, Malaysia using frequency ratio and logistic regression models. *Landslides*, 4(1): 33-41.
- Lee, S., Ryu, J.H., Won, J.S., Park, H.J. 2004. Determination and application of the weights for landslide susceptibility mapping using an artificial neural network. *Engineering Geology*, 71: 289-302.
- Leite, J.O. 1985. Interflow, overland flow and leaching of natural nutrients on an Alfisol slope of southern Bahia, Brazil. *Journal of Hydrology*, 80(1): 77-92. doi: 10.1016/0022-1694(85)90075-7.
- Liu, Z., Fan, H., Zhu, Y., & Liu, S., 2018. Prediction model for the scale of karst cave based on back propagation artificial neural network and its application. *Carsologica Sinica*, (1), 139-145.
- Ly, H.B., Thai Pham, B. 2020. Soil Unconfined Compressive Strength Prediction Using Random Forest (RF) Machine Learning Model. *The Open Construction & Building Technology Journal*, 14(1): 278-285.
- Magesh, N.S., Chandrasekar, N., Soundranayagam, J.P. 2012. Delineation of groundwater potential zones in Theni district, Tamil Nadu, using remote sensing, GIS and MIF techniques. *Geoscience Frontiers*, 3(2): 189-196.
- Malinov, S., Sha, W., & McKeown, J. J., 2001. Modelling the correlation between processing parameters and properties in titanium alloys using artificial neural network. *Computational materials science*, 21(3), 375-394.
- Mandal, S., & Mondal, S., 2019. Artificial neural network (ANN) model and landslide susceptibility. In *Statistical approaches for landslide susceptibility assessment and prediction* (pp. 123-133). Springer, Cham.
- Mannila, H. 1996. Data mining: machine learning, statistics, and databases. In: *Proceedings of 8th International Conference on Scientific and Statistical Data Base Management*, IEEE, pp.2-9.

- Marent, B.R., Portilho, S. 2017. Unidades de paisagem na bacia hidrográfica do Ribeirão Preto, Serra do Gandarela-MG. *GEOUSP Espaço e Tempo*, 21(1): 138-155.
- Masetic, Z., & Subasi, A. 2016. Congestive heart failure detection using random forest classifier. *Computer methods and programs in biomedicine*, 130: 54-64.
- Matin, S. S., Leila Farahzadi, S. Makaremi, S. Chehreh Chelgani, and G. H. Sattari. 2018. Variable selection and prediction of uniaxial compressive strength and modulus of elasticity by random forest. *Applied Soft Computing* 70: 980-987.
- Matssunaga, V. Y. Curso de Redes Neurais utilizando o MATLAB. Belém, 2012. (Apostila). Disponível em: <<http://paginas.uepa.br/engprod/index.php>> [acessado em 20 de janeiro de 2018].
- Maurity, C.W. 1995. Evolução recente da cobertura de alteração no platô N1 – Serra dos Carajas – PA. Dissertação de Mestrado, Programa de Pós-Graduação em Geologia e Geoquímica. Universidade Federal do Pará, Belém, PA.
- Mazhari, N., Shafaroudi, A.M., Ghaderi, M. 2017. Detecting and mapping different types of iron mineralization in Sangan mining region, NE Iran, using satellite image and airborne geophysical data. *Geosciences Journal*, 21(1): 137-148.
- Mohammady, M., Pourghasemi, H.R. and Amiri, M. 2019. Land subsidence susceptibility assessment using random forest machine learning algorithm. *Environmental Earth Sciences*, 78(16): 1-12.
- Mohri, M., Rostamizadeh, A., & Talwalkar, A., 2018. Foundations of machine learning. MIT press.
- Mota, N.D.O., Silva, L.V.C., Martins, F.D., Viana, P.L. 2015. Vegetação sobre sistemas ferruginosos da Serra dos Carajás. In: Carmo, F.F., Kamino, L.H.Y. Geossistemas ferruginosos do Brasil: áreas prioritárias para a conservação da diversidade geológica e biológica, patrimônio cultural e serviços ambientais. 3i Editora, Belo Horizonte, pp. 289-315.
- Mordeson J. N. Mathew S. Binu M. 2022. Applications of Mathematics of Uncertainty. Springer, 302 p.
- Myronidis, D., Papageorgiou, C., Theophanous, S. 2016. Landslide susceptibility mapping based on landslide history and analytic hierarchy process (AHP). *Natural Hazards*, 81(1): 245-263.
- Najafi, A., Karimpour, M.H., Ghaderi, M. 2014. Application of fuzzy AHP method to IOCG prospectivity mapping: A case study in Taherabad prospecting area, eastern Iran. *International Journal of Applied Earth Observation and Geoinformation*, 33: 142-154.
- National Research Council, 1996. Rock Fractures and Fluid Flow: Contemporary Understanding and Applications. Committee on Fracture Characterization and Fluid Flow. National Academy Press, Washington, DC. Accessed from: <http://www.nap.edu/catalog/2309.html>.
- Neto, A. V. C., & Baptista Filho, J. 1997. Espeleogênese em quartzitos da Serra do Ibitipoca, sudeste de Minas Gerais. *Anuário do Instituto de Geociências*, 20: 75-87.
- Ngadisih, Yatabe, R., Bhandary, N.P., Dahal, R.K. 2014. Integration of statistical and heuristic approaches for landslide risk analysis: a case of volcanic mountains in West Java Province,

- Indonesia. *Georisk: Assessment and Management of Risk for Engineered Systems and Geohazards*, 8(1): 29-47.
- Nicu, I.C. 2018. Application of analytic hierarchy process, frequency ratio, and statistical index to landslide susceptibility: an approach to endangered cultural heritage. *Environmental Earth Sciences*, 77, 79. doi:10.1007/s12665-018-7261-5
- Nobre, A.D., Cuartas, L.A., Hodnett, M., Rennó, C., Rodrigues, G., Silveira, A., Waterloo, M., Saleska, S. 2011. Height Above the Nearest Drainage - a hydrologically relevant new terrain model. *Journal of Hydrology*, 404: 13-29.
- Nobre, A.D., Cuartas, L.A., Momo, M.R., Severo, D.L., Pinheiro, A., Nobre, C.A. 2016. HAND contour: a new proxy predictor of inundation extent. *Hydrological Processes*, 30(2): 320-333.
- Oberender, P., Bauer, H., Exel, T., Lundberg, J., Scheuerer, M., & Sjöberg, R. 2014. Gobholo Cave: A granite cave in Swaziland. *International Union of Speleology Pseudokarst Commission Newsletter*, (24): 29-33.
- Oliveira, G.S.R., Jacques, P.D., Shinzato, E. 2005. Projeto APA Sul RMBH: uso e cobertura da terra, mapas cobertura e uso da terra, 1:50.000. SEMAD/CPRM, Belo Horizonte, 50 p.
- Ortiz, J. A. V., & Martínez-Graña, A. M., 2018. A neural network model applied to landslide susceptibility analysis (Capitanejo, Colombia). *Geomatics, Natural Hazards & Risk*, 9(1).
- Pahlavan-Rad, M.R., Khormali, F., Toomanian, N., Brungard, C.W., Kiani, F., Komaki, C.B., Bogaert, P. 2016. Legacy soil maps as a covariate in digital soil mapping: a case study from Northern Iran. *Geoderma*, 279: 141-148.
- Patel, J., Shah, S., Thakkar, P., & Kotecha, K. 2015. Predicting stock market index using fusion of machine learning techniques. *Expert Systems with Applications*, 42(4): 2162-2172.
- Pereira, M. C., Souza, T. 2009. A ocorrência de cavernas em minério de ferro e canga: a evolução dos estudos geoespeleológicos no Quadrilátero Ferrífero, Minas Gerais. In: *Simpósio Brasileiro de Geografia Física Aplicada, XIII, Viçosa/MG. Viçosa: UFV, 2009.*
- Pereira, M.C., Stávale, Y.O., Salgado, A.A.R. 2012. Study of the genesis of cavities and depressions on iron ore – Quadrilátero Ferrífero/MG: Rola Moça's and Gandarela's ridge. (*portuguese*) *Revista Brasileira de Geomorfologia*, 13(3): 245-253.
- Piccini, L., Mecchia, M. 2009. Solution weathering rate and origin of karst landforms and caves in the quartzite of Auyan-tepui (Gran Sabana, Venezuela). *Geomorphology*, 106: 15-25.
- Piló, L. B.; Auler, A. 2013. Introdução à Espeleologia. In: *CECAV. IV Curso de Espeleologia e Licenciamento Ambiental. Brasília, DF: Instituto Chico Mendes de Conservação da Biodiversidade*, pp.7-23.
- Piló, L.B., Auler, A.S. 2009. Geoespeleologia das cavidades em rochas ferríferas da região de Carajás, PA. In: *Rasteiro, M.A., Sarmento, R.L. (Eds.), Proceedings of the 30th Brazilian Congress of Speleology, Brazilian Speleology Society, Montes Claros*, pp.181-186.
- Piló, L.B., Auler, A.S., Martins, F. 2015a. Carajás National Forest: Iron Ore Plateaus and Caves in Southeastern Amazon. In: *Vieira, B.C., Salgado, A.A.R., Santos, L.J.C. (Eds.), Landscapes and Landforms of Brazil. Springer, Dordrecht*, pp. 273-283.
- Piló, L.B., Coelho A., Reino, J.C.R. 2015b. Geoespeleologia em rochas ferríferas: cenário atual e conservação. In: *Carmo, F.F., Kamino, L.H.Y. (Eds.), Geossistemas ferruginosos do Brasil:*

áreas prioritárias para a conservação da diversidade geológica e biológica, patrimônio cultural e serviços ambientais. 3ª Editora, Belo Horizonte, pp. 125-148.

Pipan, T., Culver, D.C. 2019. Shallow subterranean habitats. In: White, W.B., Culver, D.C., Pipan, T. (Eds.), *Encyclopedia of Caves*. Academic Press, pp. 896-908.

Pourghasemi, H.R., Pradhan, B., Gokceoglu, C. 2012. Application of fuzzy logic and analytical hierarchy process (AHP) to landslide susceptibility mapping at Haraz watershed, Iran. *Natural Hazards*, 63(2): 965-996.

Powers, D. M. (2020). Evaluation: from precision, recall and F-measure to ROC, informedness, markedness and correlation. arXiv preprint arXiv:2010.16061.

Rajabi, A. M., Khodaparast, M., & Mohammadi, M., 2021. Earthquake-induced landslide prediction using back-propagation type artificial neural network: case study in northern Iran. *Natural Hazards*, 1-16.

Rajhans, N. R., Purandare, A. S., & Pathak, S. R., 2013. Identification of parameters affecting liquefaction of fine grained soils using AHP. In *Proceedings of International Conference on Optimization procedures on engineering problems*, SVNIT, Surat.

Rasamoelina, A. D., Adjailia, F., & Sinčák, P., 2020. A review of activation function for artificial neural network. In *2020 IEEE 18th World Symposium on Applied Machine Intelligence and Informatics (SAMI)* (pp. 281-286). IEEE.

Rasyid, A.R., Bhandary, N.P., Yatabe, R. 2016. Performance of frequency ratio and logistic regression model in creating GIS based landslides susceptibility map at Lompobattang Mountain, Indonesia. *Geoenvironmental Disasters*, 3(1): 1-16.

Rennó, C.D., Nobre, A.D., Cuartas, L.A., Soares, J.V., Hodnett, M.G., Tomasella, J., Waterloo, M. 2008. HAND, a new terrain descriptor using SRTM-DEM; mapping terra-firme rainforest environments in Amazonia. *Remote Sensing of Environment*, 112: 3469-3481.

Ribas, L.M.L., Carvalho, L.C.D. 2009. Cavidade natural subterrânea: natureza jurídica. *Interações (Campo Grande)*, 10(1): 83-93.

Rocha, A. D. 2015. Estudo da assembleia de quirópteros (Mammalia: Chiroptera) da Gruta do Riacho Subterrâneo–município de Itu–SP. Dissertação de mestrado, Programa de pós-graduação em ecologia e recursos naturais. Universidade Federal de São Carlos, São Carlos, SP.

Rogers, C.M., Engelder, T. 2004. The feedback between joint-zone development and downward erosion of regularly spaced canyons in the Navajo Sandstone, Zion National Park, Utah. *Geological Society of London Special Publications*, 231(1): 49-71.

Roodposhti, M.S., Rahimi, S., Beglou, M.J. 2014. PROMETHEE II and fuzzy AHP: an enhanced GIS-based landslide susceptibility mapping. *Natural Hazards*, 73(1): 77-95.

Roy, J., & Saha, S., 2021. Integration of artificial intelligence with meta classifiers for the gully erosion susceptibility assessment in Hinglo river basin, Eastern India. *Advances in Space Research*, 67(1), 316-333.

Ruhe, R.V. 1975. *Geomorphology: geomorphic processes and surficial geology*. Houghton Mifflin, 246 p.

Russo, R. F.S.M.; Camanho, R. 2015. Criteria in AHP: a systematic review of literature. *Procedia Computer Science*, 55: 1123-1132.

- Saaty, T.L. 2004. Decision Making - the analytic hierarchy and network processes (AHP/ANP). *Journal of Systems Science and Systems Engineering*, 13(1): 1-35.
- Saaty, T.L., 1996. Decision making with dependence and feedback: The analytic network process. RWS Publications, Pittsburgh, vol. 4922.
- Sabins Jr, F.F. 1987. Remote Sensing: principles and interpretation. WH Freeman and company, New York, 506 p.
- Salgado, A.A.R., Colin, F., Nalini Jr, H. A., Braucher, R., Varajão, A.F.D.C. & Varajão, C.A.C. 2004. O Papel da Denudação Geoquímica no Processo de Erosão Diferencial no Quadrilátero Ferrífero/MG. *Revista Brasileira de Geomorfologia*, 5(1): 55-69.
- Sanchez, P.F. 2009. Mapeamento espaço-temporal e previsão de pressões piezométricas em maciços rochosos de fundações de grandes barragens – estudo de caso de Itaipu. Dissertação de mestrado, Programa de pós-graduação em construção civil. Universidade Federal do Paraná, Curitiba, PR.
- Santos, D. J. D., Ruchkys, Ú. D. A., & Travassos, L. E. P., 2021. Perfil Geocológico do Parque Nacional da Serra do Gandarela, Minas Gerais, Brasil. *Sociedade & Natureza*, 33.
- Santoso, H., Tani, H., & Wang, X. 2017. Random Forest classification model of basal stem rot disease caused by *Ganoderma boninense* in oil palm plantations. *International journal of remote sensing*, 38(16): 4683-4699.
- Sarkar, T., & Mishra, M., 2018. Soil erosion susceptibility mapping with the application of logistic regression and artificial neural network. *Journal of Geovisualization and Spatial Analysis*, 2(1), 1-17.
- Sauro, F., De Waele, J., Onac, B.P., Galli, E., Dublyansky, Y., Baldoni, E., Sanna, L. 2014. Hypogenic speleogenesis in quartzite: the case of Corona'e Sa Craba Cave (SW Sardinia, Italy). *Geomorphology*, 211: 77-88.
- Schicker, R., Moon, V. 2012. Comparison of bivariate and multivariate statistical approaches in landslide susceptibility mapping at a regional scale. *Geomorphology*, 161: 40-57.
- Segal, D. 1982. Theoretical basis for differentiation of ferric-iron bearing minerals, using Landsat MSS data. *Proceedings of Symposium for Remote Sensing of Environment, 2nd Thematic Conference on Remote Sensing for Exploratory Geology, Fort Worth*, pp.949-951.
- Senanayake, I.P., Dissanayake, D.M.D.O.K., Mayadunna, B.B., Weerasekera, W.L. 2016. An approach to delineate groundwater recharge potential sites in Ambalantota, Sri Lanka using GIS techniques. *Geoscience Frontiers*, 7(1): 115-124.
- Sepahvand, A., Ahmadi, H., Nazari Samani, A., & Trevisani, S., 2019. Landform classification of karstic area by Geomorphometric Index and Artificial Neural Network (Case study: A part of Korram Abad, Biran Shahr and Alashtar Watersheds). *Journal of Range and Watershed Management*, 72(1), 107-122.
- Sevgen, E., Kocaman, S., Nefeslioglu, H. A., & Gokceoglu, C., 2019. A novel performance assessment approach using photogrammetric techniques for landslide susceptibility mapping with logistic regression, ANN and random forest. *Sensors*, 19(18), 3940.
- Sezen, C., Bezak, N., Bai, Y., & Šraj, M., 2019. Hydrological modelling of karst catchment using lumped conceptual and data mining models. *Journal of Hydrology*, 576, 98-110.

- Shahri, A. A., Spross, J., Johansson, F., & Larsson, S., 2019. Landslide susceptibility hazard map in southwest Sweden using artificial neural network. *Catena*, 183, 104225.
- Silva, I. C. R.; Santos Junior, A. E. A. 2015. Petrografia e Caracterização de Espeleotemas em Formações Ferríferas Bandadas, Cavidade N4E-29, Serra Norte de Carajás-PA. In: Congresso Brasileiro de Espeleologia, 33. Eldorado. SBE. pp. 593-601.
- Silva, I. N.; Spatti D.; Flauzino, R. 2010. Redes neurais artificiais para engenharia e ciências aplicadas: curso prático. São Paulo: Artliber.
- Silva, L. A., Peres, S. M., & Boscarioli, C. 2016. Introdução à mineração de dados: com aplicações em R. Elsevier Brasil.
- Simmons, G.C. 1963. Canga Caves in the Quadrilátero Ferrífero, Minas Gerais, Brazil. *The National Speleological Society Bulletin*, 25: 66-72.
- Singhal, B.B.S., Gupta, R.P. 2010. Applied hydrogeology of fractured rocks. Springer Science & Business Media, London, 401 p.
- Sinha, S.K.; Wang, M.C. 2008. Artificial neural network prediction models for soil compaction and permeability. *Geotechnical Engineering Journal*, 26(1): 47-64.
- Souza, B.E.V.F. 2018. Caracterização da capacidade de armazenamento e de transmissão de água em distintos maciços de canga do Quadrilátero Ferrífero (MG). Dissertação de mestrado, Programa de pós-graduação em Geotecnia. NUGEO. Universidade Federal de Ouro Preto, Ouro Preto, MG.
- Souza, F.C.R., Carmo, F.F. 2015. Geossistemas ferruginosos no Brasil. In: Carmo, F.F., Kamino, L.H.Y. (Eds.) Geossistemas ferruginosos do Brasil: áreas prioritárias para a conservação da diversidade geológica e biológica, patrimônio cultural e serviços ambientais. 3ª Editora, Belo Horizonte, pp. 47-76.
- Souza-Silva, M., Martins, R.P., Ferreira, R.L. 2011. Cave lithology determining the structure of the invertebrate communities in the Brazilian Atlantic Rain Forest. *Biodiversity and Conservation*, 20: 1713- 1729.
- Spier, C.A., Levett, A., Rosière, C.A. 2019. Geochemistry of canga (ferricrete) and evolution of the weathering profile developed on itabirite and iron ore in the Quadrilátero Ferrífero, Minas Gerais, Brazil. *Mineralium Deposita*, 54: 983-1010.
- Srivastava, P.K., Bhattacharya, A.K. 2000. Delineation of ground water potential zones in a hard rock terrain of Bargarh district, Orissa using IRS data. *Journal of the Indian Society of Remote Sensing*, 28(3): 129.
- Stávale, Y.O. 2012. Espacialização do patrimônio espeleológico da Reserva da Biosfera da Serra do Espinhaço: geossítios selecionados e sua importância para a geoconservação. Dissertação de mestrado, Programa de pós-graduação em geografia. Universidade Federal de Minas Gerais, Belo Horizonte, MG.
- Steenhuis, T.S., Agnew, L., Gerárd-Marchant, P., Walter, M.T. 2005. Overland Flow. In: Hillel, D. (ed.). *Encyclopedia of Soils in the Environment*. Elsevier, Amsterdam, pp. 130-133.
- Summerfield, M.A. 2014. *Global Geomorphology*. Taylor & Francis Group, Routledge, 560p.
- Taheri, K., Gutiérrez, F., Mohseni, H., Raeisi, E., Taheri, M. 2015. Sinkhole susceptibility mapping using the analytical hierarchy process (AHP) and magnitude–frequency relationships: A case study in Hamadan province, Iran. *Geomorphology*, 234: 64-79.

- Tanyu, B. F., Abbaspour, A., Alimohammadlou, Y., & Tecuci, G. 2021. Landslide susceptibility analyses using Random Forest, C4. 5, and C5. 0 with balanced and unbalanced datasets. *Catena*, 203, 105355.
- Tian, Y., Xu, C., Hong, H., Zhou, Q., & Wang, D. (2019). Mapping earthquake-triggered landslide susceptibility by use of artificial neural network (ANN) models: an example of the 2013 Minxian (China) Mw 5.9 event. *Geomatics, Natural Hazards and Risk*, 10(1), 1-25.
- Trichakis, I. C., Nikolos, I. K., & Karatzas, G. P. (2011). Artificial neural network (ANN) based modeling for karstic groundwater level simulation. *Water Resources Management*, 25(4), 1143-1152.
- Twidale, C.R. 1987. Sinkholes (dolines) in lateritised sediments, western Sturt Plateau, Northern territory, Australia. *Geomorphology*, 1: 33-52.
- Tzeng, G., Huang, J. 2011. Multiple attribute decision making: methods and applications. CRC press, New York, 336 p.
- Valentim, R. F. 2016. Classificação geomecânica da Cavidade Natural subterrânea BRU_0005, município de São Gonçalo do Rio Abaixo, MG. Dissertação de mestrado, Programa de pós-graduação em Geotecnia. NUGEO. Universidade Federal de Ouro Preto, Ouro Preto, MG.
- Wang, X., Kang, Q., Qin, J., Zhang, Q., & Wang, S., 2013. Application of AHP-extenics model to safety evaluation of rock slope stability. *Journal of Central South University (Science and Technology)*, 44(6), 2455-2462.
- Wang, Y., Fang, Z., Hong, H., & Peng, L., 2020. Flood susceptibility mapping using convolutional neural network frameworks. *Journal of Hydrology*, 582, 124482.
- Wang, Z., Yi, F.C. 2009. AHP-based evaluation of occurrence easiness of geological disasters in Mianyang City. *Journal of Natural Disasters*, 18(1): 14-23.
- Webb, J.A., Finlayson, B.L. 1984. Allophane and opal speleothems from granite caves in south-east Queensland. *Journal of the Geological Society of Australia*, 31(3): 341-349.
- Weliange W.S, Dandeniya A.S, Elgiriya P, Dewage D.D, Alahakoon A.M.D, Hettiaarchchi N & P Jayasingha. 2010. Speleological Exploration to Sthreepura Lena Cave at Kuruwita, Sri Lanka. National Archaeological Symposium, 2010, pp. 1-15.
- White, W. B., Jefferson, G. L., Haman, J. F. 1966. Quartzite karst in southeastern Venezuela. *International Journal of Speleology*, 2(4): 309-314. doi:10.5038/1827-806X.2.4.1
- Wilson, J.P., Gallant, J.C. 2000. Terrain analysis: principles and applications. John Wiley & Sons, 479 p.
- Wu, Y., Jiang, X., Guan, Z., Luo, W., Wang, Y. 2018. AHP-based evaluation of the karst collapse susceptibility in Tailai Basin, Shandong Province, China. *Environmental Earth Sciences*, 77(12): 436.
- Wysocki, D.A., Schoeneberger, P.J., Hirmas, D.R., LaGarry, H.E. 2011. Geomorphology of soil landscapes. In: Huang, P.M., Li, Y., Sumner, M.E. (Eds.). *Handbook of soil sciences: properties and processes*. CRC Press, Boca Raton, 29, pp. 1-26.
- Xia, Y. and Wu, C. 2014. A Neural Network Method to Calculate the Pre-consolidation Stress of Soil. *Electronic Journal of Geotechnical Engineering*, 19(Z7): 17797-17804.

- Xie, P., Wen, H., Ma, C., Baise, L. G., & Zhang, J., 2018. Application and comparison of logistic regression model and neural network model in earthquake-induced landslides susceptibility mapping at mountainous region, China. *Geomatics, Natural Hazards and Risk*, 9(1), 501-523.
- Yang, X., Ding, J., Hou, H. 2013. Application of a triangular fuzzy AHP approach for flood risk evaluation and response measures analysis. *Natural Hazards*, 68(2): 657-674.
- Zadeh, L.A. 1965. Fuzzy sets. *Information and control*, 8(3): 338-353.
- Zhang, J., Huang, Y.F., Munasinghe, D., Fang, Z., Tsang, Y.P., Cohen, S. 2018. Comparative analysis of inundation mapping approaches for the 2016 flood in the Brazos River, Texas. *Journal of the American Water Resources Association*, 54(4): 820-833.
- Zhang, N., Zhou, K., Du, X. 2017. Application of fuzzy logic and fuzzy AHP to mineral prospectivity mapping of porphyry and hydrothermal vein copper deposits in the Dananhu-Tousuquan island arc, Xinjiang, NW China. *Journal of African Earth Sciences*, 128: 84-96.
- Zhu, J. and Pierskalla Jr, W.P. 2016. Applying a weighted random forests method to extract karst sinkholes from LiDAR data. *Journal of Hydrology*, 533: 343-352.

DECLARAÇÃO DE INEXISTÊNCIA DE PLÁGIO

Eu IRAYDES TÁLITA DE SENA NOLA, declaro que, a tese com título " ANÁLISE MULTICRITÉRIO E APRENDIZADO DE MÁQUINA APLICADOS NA PREDIÇÃO DO POTENCIAL ESPELEOLÓGICO DA REGIÃO DO PARQUE NACIONAL SERRA DO GANDARELA, QUADRILÁTERO FERRÍFERO/MG", com exceção das citações diretas e indiretas claramente indicadas e referenciadas, foi escrita por mim e, portanto, não contém plágio, tal como definidos pela legislação de direitos autorais em vigor no Brasil, Lei 9.610, de 19 de fevereiro de 1998. Eu estou consciente que a utilização de material de terceiros incluindo uso de paráfrase sem a devida indicação das fontes será considerado plágio, e estarei sujeito processo administrativo e outras sanções legais.

Ouro Preto, 10 de agosto de 2022

