



Universidade de Brasília

Instituto de Geociências

Programa de Pós-Graduação em Geociências Aplicadas e Geodinâmica

**APLICAÇÃO DE *RANDOM FOREST* PARA A DETECÇÃO
DE CORPOS MÁFICOS-ULTRAMÁFICOS DA SUÍTE
GAMELEIRA NO EMBASAMENTO DA FAIXA
BRASILIA, BRASIL**

Luiz Fernando Cursino Tabosa

Orientador: Elder Yokoyama

Dissertação de mestrado N° 222

Brasília, Setembro de 2023

Luiz Fernando Cursino Tabosa

**APLICAÇÃO DE *RANDOM FOREST* PARA A DETECÇÃO
DE CORPOS MÁFICOS-ULTRAMÁFICOS DA SUÍTE
GAMELEIRA NO EMBASAMENTO DA FAIXA
BRASILIA, BRASIL.**

Dissertação submetida ao Programa de
Pós-Graduação em Geociências Aplicadas
e Geodinâmica da Universidade de
Brasília como requisito para obtenção do
Título de Mestre em Geofísica Aplicada

Orientador: Prof. Dr. Elder Yokoyama

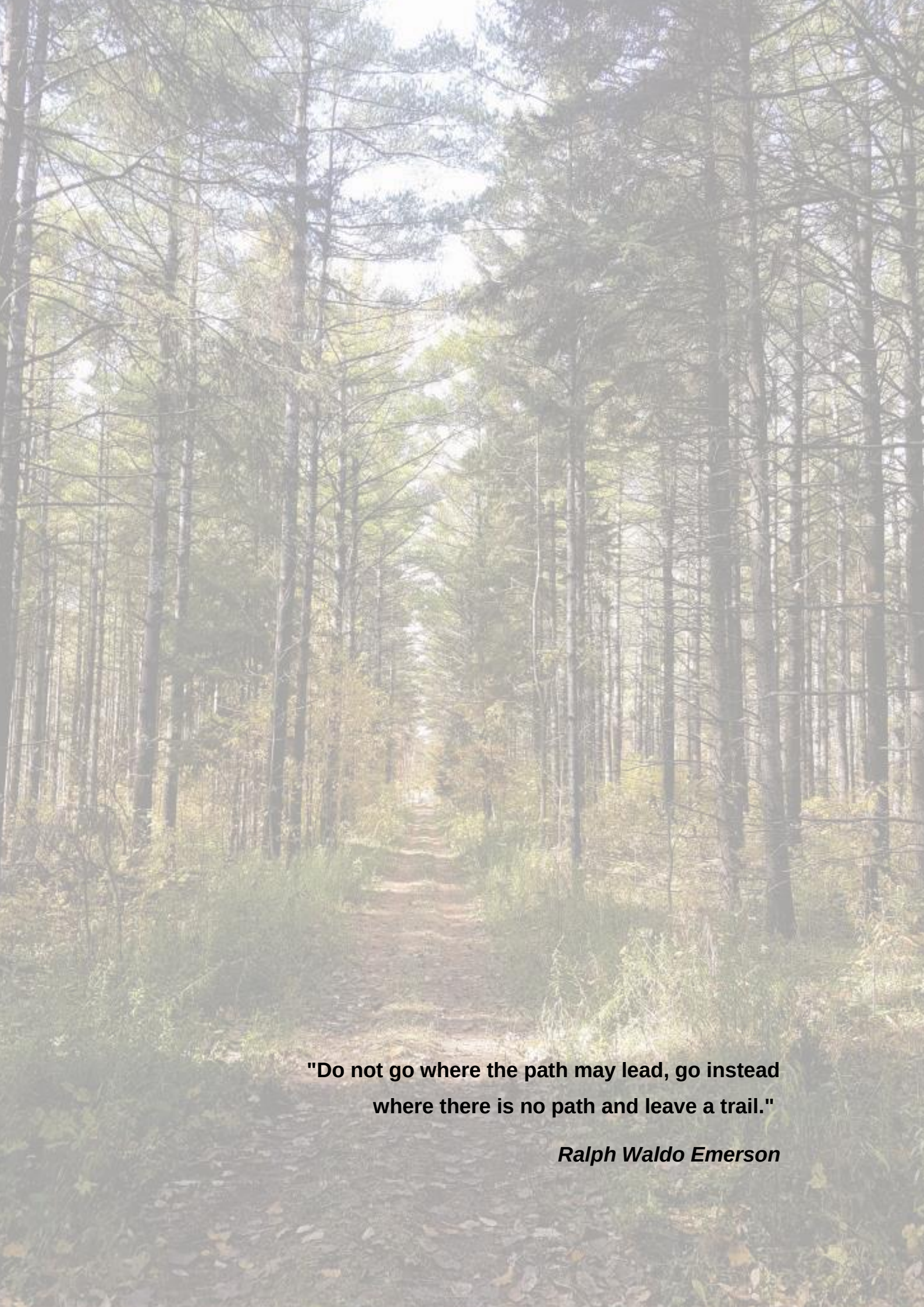
Brasília, setembro de 2023

Ficha catalográfica elaborada automaticamente,
com os dados fornecidos pelo(a) autor(a)

C Cursino Tabosa, Luiz Fernando
 APLICAÇÃO DE RANDOM FOREST PARA A DETECÇÃO DE CORPOS
 MÁFICOS-ULTRAMÁFICOS DA SUÍTE GAMELEIRA NO EMBASAMENTO DA
 FAIXA BRASÍLIA, BRASIL / Luiz Fernando Cursino Tabosa;
 orientador Elder Yokoyama. -- Brasília, 2023.
 109 p.

 Dissertação (Mestrado em Geociências Aplicadas) --
 Universidade de Brasília, 2023.

 1. Aerogeofísica. 2. Faixa Brasília. 3. , Machine
 learning. 4. Random Forest. 5. Rochas Máficas Ultramáficas.
 I. Yokoyama, Elder, orient. II. Título.



**"Do not go where the path may lead, go instead
where there is no path and leave a trail."**

Ralph Waldo Emerson

Agradecimentos

Gostaria de agradecer primeiramente a oportunidade que a vida me deu de poder realizar um mestrado, acredito que esse seja um privilégio dentro do cenário brasileiro. Agradeço a todas as entidades que me protegem, e que foram responsáveis por me motivarem e me guiarem muitas vezes em momentos difíceis.

A minha família, que como sempre agem como uma fonte de suporte infinito, amo todos incondicionalmente. A meus pais, Iara e Pena, que conseguiram em meio a muita luta me dar instrução para batalhar todos os dias para alcançar meus sonhos. Meu irmão Di, que de seu jeito sempre esteve do meu lado e sem nenhuma dúvida um dos maiores suportes que tenho. Meus irmãos, Felipe e Rodrigo que sempre estiveram me acompanhando e torcendo por mim. A minha mana, Lulis, meu porto seguro, minha grande amiga que sempre sabe o que falar seja em algumas horas palavras de conforto e em outras o que realmente precisava ouvir. A minha tia e minha avó, Claudia e Antônia, que me ensinaram a ser humilde e sempre me levantar independente de qualquer coisa. Sem vocês nada disso faria sentido.

A minha companheira de vida Julia, pelo carinho, amor, incentivo e por cotidianamente me lembrar que consigo alcançar lugares que nem imaginei ainda. Obrigado por compartilhar os dias comigo, e me permitir ver um pouco do mundo sob sua perspectiva.

Gostaria de agradecer a meu orientador, amigo, Elder Yokoyama que sempre acreditou em mim, arrisco dizer que em vezes mais do que eu mesmo acreditei em mim. Obrigado pelos ensinamentos que não se restringiram somente as geociências, mas muito aprendi sobre a vida.

A minha equipe incrível de campo, Pedro Maragno e Mendes, vocês foram sensacionais. Toparam ir para uma área remota, ficar acampado, e todos os dias sempre estiveram com um sorriso na cara independente das dificuldades, vocês transformaram algo extremamente exigente em algo leve e prazeroso.

Meu irmão de vida Thiago Mendes, que por coincidência entramos no mestrado ao mesmo tempo e dividimos um objeto de pesquisa semelhante. Agradeço a tua paciência, parceria, teus conselhos e nossas inúmeras conversas com toda certeza tiveram uma grande influência nesse momento.

Meus colegas de sala, Boi e Vagney, que tiveram um papel muitas vezes de mentores de projetos, ciência e questões de vida, agradeço as infindáveis conversas e todos os ensinamentos,

Gostaria de agradecer a meus amigos de vida que me apoiaram e acreditaram em mim, Curu, Thed, Liam, Decio, Pellicano, Dadu, Toni, Aninha, Zaka, Bela, Gaby, Lucatelli, Jetulio, Kaverna, Vitinho Goodadventure, Nena. E a todos que acompanharam essa jornada!

Ao fomento disponibilizado pela CAPES. O investimento na educação superior no Brasil é fundamental para o desenvolvimento pleno da nossa sociedade, e o apoio financeiro fornecido pela CAPES permite a realização de pesquisas como essa. O presente trabalho foi realizado com o apoio da Coordenação de Aperfeiçoamento de Pessoal de Nível Superior- Brasil (CAPES) - Código de Financiamento 001.

No mais, keep Choosing life, fight the system and stay true to yourself!

Resumo

O embasamento do Orógeno Brasília em sua porção norte, corresponde a parte de um bloco crustal gerado no início do Paleoproterozoico, sendo formado por um extenso volume de magmatismos TTG (trondhjemitó-tonalito-granodiorito). A Suíte Gameleira, ocorre no embasamento da Faixa Brasília Norte (sudeste do estado do Tocantins), e corresponde a corpos predominantemente máficos com ocorrência subordinada de rochas ultramáficas no Domínio Almas-Conceição do Tocantins. A região é composta por variações de granitoides, sequências metavulcanossedimentares e metassedimentares. Dessa maneira, o objetivo dessa pesquisa é usar o algoritmo de aprendizado de máquina *Random Forest* para a detecção de ocorrências de novos corpos máfico-ultramáficos na Suíte Gameleira. Para isso, foram utilizados dados remotos, como levantamentos aerogeofísicos (magnetometria e radiometria), além de sensoriamento remoto (Alos Palsar). Foram gerados produtos derivados da aerogeofísica como Intensidade Magnética Total, Amplitude do Sinal Analítico, uma fatia em -100 metros de profundidade do resultado da Inversão dos Vetores de Magnetização, derivada vertical, derivada Tilt, os três canais radiométricos (K, eTh e eU), Contagem Total, Fator F e Potássio Anômalo. Com a informação desses produtos, foi gerada a base de dados e aplicado o algoritmo *Random Forest* (algoritmo de aprendizado de máquina supervisionado). A base de dados de treino possui 1659 amostras equivalente a 1,96% da base de dados total (84535 amostras) para a geração de mapas preditivos litológicos. Para a verificação e avaliação dos dados preditos foram usadas três abordagens: uso de lineamentos magnéticos para averiguar um controle tectônico estrutural na colocação dos corpos, amostras de campo e o uso da Inversão dos Vetores de Magnetização para a análise do comportamento magnético em profundidade dos corpos preditos, uma vez que os corpos da Suíte Gameleira apresentam remanência magnética. Concluiu-se na pesquisa que o uso desse método atrelado as técnicas de verificação utilizadas auxiliam na exploração de terrenos *greenfield* e *brownfield*, uma vez que o aprendizado de máquinas pode aprimorar a cartografia existente, delimitando novas ocorrências ou mesmo sugerindo áreas em que se faz necessário melhor detalhamento. O uso de conhecimentos geológicos iniciais e outras informações não inseridas no modelo são de extrema importância, uma vez que o mapa preditivo não corresponde em si a verdade geológica e feita a predição é necessária que seja realizada interpretação do resultado por uma visão geológica.

Palavras-chave: Aerogeofísica, Faixa Brasília, Machine learning, Random Forest, Rochas Máficas Ultramáficas

Abstract

The northern part of the Brasília Orogen in its basement corresponds to a portion of a crustal block formed in the early Paleoproterozoic and consists of an extensive TTG magmatism (trondhjemite-tonalite-granodiorite). The Gameleira Suite is found in the basement of the North Brasília Belt (southeastern Tocantins) and consists of predominantly mafic rocks with subordinate occurrences of ultramafic rocks in the Almas-Conceição do Tocantins Domain. The region is composed of variations of granitoids, metavolcanosedimentary sequences, and metasedimentary rocks. The research aims to use a machine learning method for detecting potential new mafic-ultramafic occurrences in the Gameleira Suite. To achieve this, remote sensing data such as aerogeophysical surveys (magnetometry and radiometry) were used, along with remote sensing. Derived products from aerogeophysics were generated, including Total Magnetic Intensity, Analytic Signal Amplitude, a slice at -100 meters depth from the Magnetic Vector Inversion, vertical derivative, Tilt derivative, the three radiometric channels (K, eTh, and eU), Total Count, F-factor, and Potassium Anomaly. Using the information from these products, a database was created and the Random Forest algorithm (a supervised machine learning algorithm) was applied, training on 1.96% of the total database (84,535 samples) to generate the predictive map. Three approaches were used for data verification and evaluation: the use of magnetic lineaments to investigate structural tectonic control on the placement of bodies, field samples, and the use of MVI to analyze the magnetic behavior at depth of the predicted bodies, as the Gameleira Suite bodies exhibit magnetic remanence. The research concluded that the use of this method, combined with the verification methods used, assists in greenfield and brownfield exploration. Machine learning can enhance existing cartography by delineating new occurrences or suggesting areas that require further detailed mapping. The use of initial geological knowledge and other information not included in the model is crucial since the predictive map does not inherently represent geological truth, and interpretation of the results through a geological perspective is necessary.

Keywords: Aerogeophysics, Brazilian Belt, Machine learning, Random Forest, Ultramafic Mafic Rocks

Sumário

1. Introdução	1
1.1 Apresentação	1
1.2 Objetivo Geral	3
1.2.1 Objetivos Específicos	3
1.3 Localização da área de estudo	3
1.4 Estrutura da Dissertação	5
2. Geologia Regional	6
2.1 Faixa Brasília	6
2.2. Bloco Crustal Cavalcante-Natividade (BCCN)	8
2.2.1 Embasamento Paleoproterozoico - Domínio Almas-Conceição do Tocantins (DACT)	8
3. Aprendizado de Máquinas	18
3.1 Random Forest	21
3.2 Conceitos de Aprendizado de Máquinas	22
3.2.1 Overfitting e underfitting	22
3.2.2 <i>Area Under the ROC Curve</i> (AUC)	23
3.2.3 Acurácia (<i>Accuracy</i>)	24
3.2.4 Precisão (<i>Precision</i>)	25
3.2.5 <i>Recall</i>	25
3.2.6 F1-Score	25
3.2.7 Validação Cruzada	25
4. Materiais e Métodos	27
4.1. Aerogeofísica	27
4.1.1 Características do Projeto Aerogeofísico Tocantins	27
4.1.2 Gamaespectrometria	28
4.1.2.1. Processamento de Dados Gamaespectrométricos – Técnicas de realce	30
4.1.3 Magnetometria	34
4.1.3.1 Processamento de Dados Magnetométricos – Técnicas de Realce	36

4.2 Criação e elaboração da Base de dados para implementação do algoritmo <i>Random Forest</i>.	43
4.3 Random Forest	45
Capítulo 5 – Resultados	47
<u>Artigo</u> Uso de <i>Random Forest</i> para a Detecção de Rochas Máficas e Ultramáficas da Suíte Gameleira, no Embasamento Arqueano Paleoproterozoico da Faixa Brasília Norte.	47
1. Introdução	49
2. Geologia Regional	50
3. Materiais e Métodos	52
3.1 Dados Litológicos	52
3.2 Dados Aerogeofísicos	54
4. Implementação do Random Forest	59
4.1 Pré-processamento e elaboração da base de dados	59
4.2 <i>Random Forest</i>	61
5. Resultados	63
5.1 Dados de campo	63
5.2 Análise exploratória dos dados	64
5.3 Mapas de classificação utilizando Random Forest	68
5.4 Lineamentos magnéticos e predições geradas para a Suíte Gameleira	73
5.5 Inversões do Vetor de magnetização e posicionamento de corpos máficos ultramáficos	74
6. Discussões	75
6.1 Predições e classificação	76
6.2 Verificação indireta ou remota	78
7. Conclusões	79
8. Referências Bibliograficas	Erro! Indicador não definido.
Capítulo 6 – Considerações Finais	81
6.1 Sugestões para trabalhos futuros	82

Índice de Figuras

Figura 1: Mapa de localização indicando as principais rodovias e estradas de acesso, cidades, e redes drenagem inseridas na área de estudo. Em destaque em azul para os corpos já cartografados da Suíte Gameleira. 4

Figura 2: Mapa geológico esquemático do terreno Almas-Dianópolis com idades U-Pb em evidência e delimitando os diferentes domínios tectônicos (Adaptado de Fuck et al., 2014). O retângulo em destaque vermelho indica a área de estudo inserida no domínio 5, a unidade nomeada Greenstone Belt no mapa corresponde ao Grupo Riachão do Ouro e a Suíte Gameleira. 7

Figura 3: Mapa geológico da área de estudo, modificado de Saboia, 2021; Saboia *et al.*, 2019; Abdalah *et al.*, 2015; SGB, 2022. 10

Figura 4: Resultados do Projeto: “Geologia e resultados prospectivos da Área Barra do Gameleira – Tocantins” (CPRM, 2000). A - Mapa geológico da intrusão Barra do Gameleira, predominância de rochas máficas e subordinamente ultramáficas controlados por estruturas. *Link para download:*
<<https://rigeo.sgb.gov.br/bitstream/doc/1610/1/Geologia%20e%20resultados%20prospectivos%20da%20%20c3%a1rea%20Barra%20da%20Gameleira%20-%20Tocantins.pdf>> 15

Figura 5: Representação hierárquica dos diferentes conceitos da inteligência artificial. Fonte: <<https://opencadd.com.br/machine-learning-ou-deep-learning/>>, acessado em 17/04/2023. 18

Figura 6: Processo de decisão de classe pelo algoritmo Random Forest. Fonte: <<https://www.tibco.com/pt-br/reference-center/what-is-a-random-forest>>, acessado em 17/04/2023. 21

Figura 7: Exemplos de formas de curvas a depender do ajuste do dado. A - *Underfitting* para um modelo linear; B - Bom ajuste para um modelo quadrático; C - *Overfitting* para um modelo polinomial de grau 15 (Adaptado de Burkov, 2019). 23

Figura 8: Exemplos de curvas para diferentes valores de AUC abaixo da ROC. Extraído de Burkov, 2019. 24

Figura 9: Localização da área do levantamento do Projeto Aerogeofísico Tocantins, em vermelho e a área de magnetometria e gamaespectrometria dos Levantamentos

Aerogeofísicos da Bacia do Parnaíba em verde. Tabela contendo as coordenadas dos vértices do levantamento. Modificado de CPRM, 2006. 28

Figura 10: Tabela com teores de concentração de K, U e Th das principais rochas e solos da Austrália. (Adaptado de Dickson e Scott, 1997). 30

Figura 11: Imagens geradas a partir do processamento dos dados radiométricos, imagens sombreadas: A: Contagem Total (CT); B: Concentração de Potássio (K); C: Concentração de Tório Equivalente (eTh); D: Concentração de Urânio Equivalente (eU). 31

Figura 12: Imagens geradas a partir do processamento dos dados radiométricos, imagens sombreadas: A – Composição Ternária (Red: K, Green: eTh, Blue: eU); B – Fator F (%); C – Potássio Anômalo; D – Índice Máfico. 35

Figura 13: Intervalos de susceptibilidade magnética para diferentes minerais e tipos de rocha. As porções de coloração mais forte indicam os intervalos mais comuns. Retirado e modificado de Dentith & Mudge, 2014. 37

Figura 14: Imagens geradas a partir do processamento dos dados magnetométricos, imagens sombreadas: A: Total Magnetic Intensity (TMI); B: Amplitude do Sinal Analítico (ASA); C: Gradiente Horizontal Total (GHT); D: Derivada Vertical (DZ). 40

Figura 15: Comparação dos métodos de inversão de susceptibilidade e MVI, com valores de susceptibilidade equivalentes. Em azul, inversão convencional de susceptibilidade e em rosa, inversão dos vetores de magnetização (MVI). O MVI apresenta uma melhor centralização das anomalias, enquanto a inversão por susceptibilidade traz uma geometria mais alongada das anomalias. 43

Figura 16: Diagrama indicando a proporção do número de amostras para cada uma das 17 classes litológicas. As cores das barras correspondem as mesmas cores do mapa geológico. O código das unidades se encontra na lista de abreviações, no início do trabalho. 44

Figura 17: Fluxograma esquemático do processamento de dados aerogeofísico e composição da base de dados utilizada para implementação do Random Forest. 45

Fig. 1. Mapa litológico extrapolado da área de estudo, modificado de Saboia, 2021; Saboia *et al.*, 2019; Abdalah *et al.*, 2015; Ribeiro *et al.*, 2022. 53

Fig. 2. Produtos aeroradiométricos gerados para a área de estudo. A – Contagem Total; B – Canal do Potássio; C – Canal do Tório; D – Canal do Urânio; E – Fator F; F - Potássio Anômalo. 57

Fig. 3. Produtos aeromagnéticos gerados para a área de estudo. A – Intensidade Magnética Total; B – Amplitude do Sinal Analítico; C – Gradiente Horizontal Total; D – *Magnetic Vectors Inversion* – slice de -100 metros de profundidade; E – Índice Máfico; F – Altimetria. 60

Fig. 4. A - Localização dos pontos de treinamento utilizado nos experimentos para implementação do Random Forest. A base de dados de treino é composta por 1659 pontos, sendo 100 amostras aleatórias de cada classe, e 59 pontos da classe dos diques cenozoicos; B – Diagrama indicando a proporção do número de amostras para cada uma das 17 classes litológicas. 62

Fig. 5. Pontos de visitação em campo e litotipos descritos. A e G – Hornblenda Gabro; B e H – Olivina Gabro Norito; C -Dunito Serpentinizados; D – Gabro Norito; E e F - Leuco Gabros. Em vermelho o contorno das unidades cartografadas no mapa litológico. 64

Fig. 6. Diagrama Quantil-Quantil (QQ-plot) para os parâmetros físicos utilizados. Em verde a distribuição dos pontos, a linha preta indica o comportamento esperado para uma distribuição normal. 65

Fig. 7. Diagramas do tipo Box Plot, elaborado para caracterização e visualização da assinatura de cada unidade, classe, com relação aos parâmetros físicos K, eTh, eU, ASA, respectivamente, A, B, C e D. O eixo X corresponde as seguintes unidades: Coberturas detrito-lateríticas (Cdl), Depósitos aluvionares (DepA), Diques (Dique), Gr. Bambui (GrBam), St. Gameleira (SGm), St. Granodiorítica aTonalítica (SGT), Un. Monzogranítica (UnMonz), Gr. Natividade (GrNat), St. Peraluminosa (SP), St. Peraluminosa-tonalítica (SPT), Suíte Quartzo-Diorítica (SQD), St. Ribeirão Itaboca (RIS), St. Ribeirão das Areias (RAC), St. Ribeirão das Areias (RAC-IB), St. Serra Boqueirão (SSB), Fm. Morro do Carneiro (FMC) e Fm. Córrego do Paiol (FCP). 67

Fig. 8. Mapas resultantes da classificação litológica realizada pelo algoritmo Random Forest com base nos três canais radiométricos (K, eTh e eU), altimetria (Alos) diferentes entradas de dados magnéticos. A – Teste A-01 (TMI, K, eTh, eU e MDT); B - Teste A-03 (ASA, K, eTh, eU e MDT); C - Teste A-05 (MVI, K, eTh, eU e MDT); D - Teste A-07 (TMI, ASA, MVI, K, eTh, eU e MDT). 69

Fig. 9. Mapas resultantes do Teste A-09, classificação litológica realizada pelo algoritmo Random Forest com base nos parâmetros: TMI, ASA, MVI, K, eTh, eU, Kd, Alos e IM. Em A – Resultado bruto da classificação; B – Mapa probabilístico indicando em cores

quentes e frias, altas e baixas probabilidades, respectivamente; C – Mapa preditivo litológico com uma filtragem por Majority; D – Mapa litológico de referência. 70

Fig. 10. Gráficos indicando o número total de amostras classificadas de maneira distinta da classe original, e o percentual de cada classe classificada no lugar. 72

Fig. 11. Mapas resultantes da classificação litológica realizada pelo algoritmo Random Forest com base nos parâmetros: A – Teste A-02 (TMI, K, eTh, eU, Alos e X_UTM); B – Teste A-10 (TMI, ASA, MVI, K, eTh, eU, Kd, Alos, IM e X_UTM). 73

Fig. 12. Mapa indicando lineamentos magnéticos extraídos dos produtos Dz e TILT, em preto. Em escala de cor de azul a vermelho temos a probabilidade de ocorrência de corpos pertencentes a Suíte Gameleira. 74

Fig. 13. Comparativo entre os produtos da inversão dos vetores de magnetização (MVI) e o produto do Random Forest referente a probabilidade de ocorrência da Suíte Gameleira, os polígonos em preto correspondem aos limites gerados pelo algoritmo. A – Produto referente a área de estudo como um todo, a variação de cinza a branco indica a menor para a maior probabilidade de ocorrência da unidade; B, C, D e E – Produtos referentes as subáreas destacadas em A, que apresentam maior probabilidade de ocorrência. Na base do modelo encontra-se o TMI. 75

Índice de Tabelas

Tabela 1: Síntese das 17 classes litológicas da área de estudo. Informações retiradas de Cruz & Kuyumjian, 1998, Kuyumjian *et al.*, 2012, Abdallah & Meneghini, 2017, Saboia & Meneghini, 2019, Saboia, 2021. 52

Tabela 2: Informação de parâmetros utilizados na realização de testes e suas métricas resultantes. 68

Tabela 3: Matriz de confusão comparando classes mapeadas e predições do *Random Forest*. Os valores indicam a percentagem de amostras mapeadas presentes no mapa preditivo. Os destaques em verde, azul, amarelo, vermelho e cinza indicam um *recall** maior que 80%, 70%, 60%, 50% e inferior a 50%, respectivamente. O valor total no eixo X corresponde ao total de pontos que foi predito para a unidade, já o total no eixo Y o número de amostras de teste para a unidade. 71

Lista de Abreviaturas e Siglas

AGP-LA - AeroGeoPhysica Latino América
ANP - Agência Nacional do Petróleo
ASA - Amplitude do Sinal Análítico
AUC – Área Abaixo da Curva (*Area Under the Curve*)
BCCN - Bloco Crustal Cavalcante-Natividade
CART - *Classification and Regression Tree*
CMA – Campo Magnético Anômalo
CMT - Campo Magnético Total
CSF - Cráton São Francisco
DACT - Domínio Almas-Conceição do Tocantins
Dx, Dy, Dz - Derivadas Direcionais em X, Y e Z
FB - Faixa Brasília
FBN - Faixa Brasília Norte
FBS - Faixa Brasília Sul
GHT - Gradiente Horizontal Total
GRO - Grupo Riachão do Ouro
IBG – Intrusão Barra do Gameleira
IGRF - *International Geomagnetic Reference Field*
Kd – Potássio Anômalo
Ki – Potássio Ideal
Ku – Urânio Anômalo
MMA - Ministério do Meio Ambiente
MPM - Mapas de Prospectividade Mineral
MVI – Inversão dos Vetores de Magnetização (*Magnetic Vectors Inversion*)
NASA - *National Aeronautics and Space Administration*
Nt - Nanotesla
OMS - Organização Mundial da Saúde
PGE – Elementos do Grupo da Platina (*Platinum Group Elements*)
REE – Elementos Terras Raras (*Rare Earth Elements*)
ROC - *Receiver Operating Characteristic*

SBG-CPRM - Serviço Geológico do Brasil

TFP - Taxa de Falsos Positivos

TPR - Taxa de Verdadeiros Positivos

TMI – Intensidade Magnética Total (*Total Magnetic Intensity*)

Ui – Urânio Ideal

Siglas de Unidades Geológicas

Coberturas detrítico-lateríticas	Cdl
Depósitos aluvionares	DepA
Diques	Diques
Gr. Bambui	GrBam
St. Gameleira	SGm
St. Granodiorítica aTonalítica	SGT
Unit Monzogranítica	UnMonz
Gr. Natividade	GrNat
St. Peraluminosa	SP
St. Peraluminosa–tonalítica	SPT
Suíte Quartzo-Diorítica	SQD
St. Ribeirão Itaboca	RIS
St. Ribeirão das Areias	RAC
St. Ribeirão das Areias	RAC-IB
St. Serra Boqueirão	SSB
Fm. Morro do Carneiro	FMC
Fm. Córrego do Paiol -	FCP

1. Introdução

1.1 Apresentação

As ferramentas utilizadas para o mapeamento geológico vêm se atualizando de forma que o uso de técnicas indiretas para a aquisição de informações, na etapa pré-campo, tem se tornado cada vez mais comuns e essenciais. O Brasil possui extensa área recoberta por aerolevantamentos radiométricos e aeromagnetométricos de boa resolução espacial, disponibilizados de maneira pública, através do Serviço Geológico Brasileiro (SGB-CPRM), de forma que, o uso desse tipo de dado, no início de uma campanha prospectiva, tem se tornado essencial.

Nessa linha de pensamento, com o advento da mineração de dados e a evolução dos métodos estatísticos, o emprego de aprendizado de máquina (*Machine Learning*) vem se tornando uma ferramenta cada vez mais utilizada no cenário prospectivo e na compreensão de ambientes geológicos (Schnitzler *et al.*, 2019). De fato, a utilização dessas ferramentas tem ganhado cada vez mais importância na atividade mineraria mineração, devido ao seu caráter de baixo custo e ampla cobertura espacial que possibilita a avaliação de grandes áreas em busca de ocorrências minerais (Costa *et al.*, 2020, Hill *et al.*, 2021). Nesse cenário, a busca por minerais críticos que contenham elementos como cobre, lítio, cromo, cobalto, elementos do grupo da platina (PGE), elementos terras-raras (REE) por exemplo, tem sido cada vez mais latente devido à grande demanda de indústrias tecnológicas (automóveis elétricos, telecomunicação, energia renovável entre outras).

Hoje, a sociedade enfrenta as consequências das mudanças climáticas, impulsionando acordos globais para uma economia com baixa emissão de carbono (Acordo de Paris COP21, 2015). Alguns países planejam até 2050 possuir uma economia livre da emissão de carbono (EU, 2020). Essa transição, tem por consequência o aumento significativo na demanda desses minerais críticos.

Quanto ao contexto geológico, os depósitos que alojam essas commodities possuem sistemas minerais diversos, onde cada sistema pode abranger um grupo de elementos diferentes, como o sistema mineral máfico-ultramáfico ortomagmático, por exemplo, gerador de depósitos sulfetados de Ni-Cu, Ni-Cu-Co e Ni-Cu-PGE (Skirrow *et al.*, 2013). Alguns dos maiores depósitos minerais de classe mundial relacionados a intrusões máfica-ultramáficas conhecidos são: Norilsk-Talnakh (e. g. Naldrett, 2004, Barnes, 2020)

e Pechenga (e. g. Naldrett, 2004) na Rússia, Bushveld (e. g. Naldrett, 2004, Cawthorn, 2010) na África do Sul, Voisey's Bay (Naldrett, 1999, e.g. 2004), Sudbury (e. g. Naldrett, 2004, Dare *et al.*, 2011) e Thompson (Naldrett, 1998) no Canadá, Jinchuan (Chai, 1992, Naldrett, 1999) na China. No Brasil destacam-se alguns depósitos minerais hospedados em rochas máfica-ultramáficas como, por exemplo: Complexo Máfico Ultramáfico Luanga (Mansur & Ferreira Filho, 2016, Mansur *et al.*, 2020), Americana do Brasil (Silva, 2009, Silva *et al.*, 2011) os três complexos máfico-ultramáficos localizados no arco magmático do Goiás Niquelândia (Ferreira-Filho *et al.*, 2010, Barbosa, 2013), Cana Brava (Ferreira Filho, 2010, Giovanardi, 2017) e Barro Alto (Ferreira Filho, 2010, Giovanardi, 2017).

A busca por novas jazidas e corpos que possam ser prospectados de forma economicamente viável se torna cada vez mais necessária. Este trabalho tem o foco no uso de ferramentas para a prospecção de novos corpos máfico-ultramáficos pertencentes à Suíte Gameleira (Correia Filho & Sá, 1980, CPRM, 2000, Saboia, 2021), no sudeste do Tocantins, entre os municípios de Conceição do Tocantins, Dianópolis e Almas. Até então, foram cartografados em torno de 12 corpos pertencentes à Suíte Gameleira (Abdalah *et al.*, 2015, Saboia *et al.*, 2019, Ribeiro *et al.*, 2022) com tamanhos variados, com média de 8 km², e formatos semi-circulares (Figura 1). A maior intrusão pertencente a Suíte Gameleira foi estudada com viés prospectivo pela CPRM, em 2000, onde foi realizado mapeamento geológico, análise de sedimentos de corrente e concentrados de bateia, que indicaram concentrações razoáveis de Ni-Cu-Cr além de paládio, elemento pertencente aos Elementos do Grupo da Platina.

Os estudos prospectivos focados na Suíte Gameleira são poucos, sendo eles: "Geologia e resultados prospectivos da Área Barra do Gameleira – Tocantins" (CPRM, 2000) e "Mapeamento da Assinatura de Ni-Cu (Co-Au-PGE) na Província Tocantins: o Exemplo da Intrusão Máfica-Ultramáfica Barra do Gameleira" (Mendes, 2023), de forma que o potencial real desta unidade ainda não é claro.

Em vista disto, o presente trabalho visa contribuir com a localização de novos corpos, que podem ser foco de futuras pesquisas prospectivas ou mesmo no avanço da compreensão da gênese dessa unidade, além do avanço no conhecimento referente ao uso de aprendizado de máquina em geociências.

1.2 Objetivo Geral

O principal objetivo dessa dissertação de mestrado é localizar novas ocorrências de corpos máfico-ultramáficos da Suíte Gameleira (TO) por meio de mapas preditivos gerados pelo algoritmo *Random Forest*.

1.2.1 Objetivos Específicos

- i. Confeccionar produtos aerogeofísicos individualizando anomalias magnéticas e radiométricas as associando aos corpos máfico-ultramáficos de interesse;
- ii. Analisar de maneira estatística os parâmetros físicos adquiridos por aerogeofísica para cada unidade da área de estudo;
- iii. Realizar testes com o algoritmo *Random Forest*, utilizando diferentes parâmetros do algoritmo, bem como diferentes tipos de dados de entrada (produtos aeromagnetométricos, aeroradiométricos e de sensoriamento remoto);
- iv. Elaborar modelos de inversão com o uso do algoritmo *Magnetization Vector Inversion* (MVI);
- v. Relacionar o resultado dos modelos de inversão como uma estratégia de verificação para os modelos litológicos gerados a partir do aprendizado de máquinas;

1.3 Localização da área de estudo

A área de estudo está localizada no sudeste do estado do Tocantins e abrange os municípios de Almas, Conceição do Tocantins, Dianópolis, Porto Alegre do Tocantins e Taipas do Tocantins (Figura 1). As principais rodovias de acesso à área de estudo correspondem às rodovias TO-050, TO-485, TO-387 e TO-040.

A coleta de amostras para o estudo da litologia das intrusões pertencentes a Suíte Gameleira teve seu acesso em parte por rodovias, no entanto, as ocorrências na região central da área de estudo foram acessadas por estradas de terra e possuíam um difícil acesso para realização de perfis.

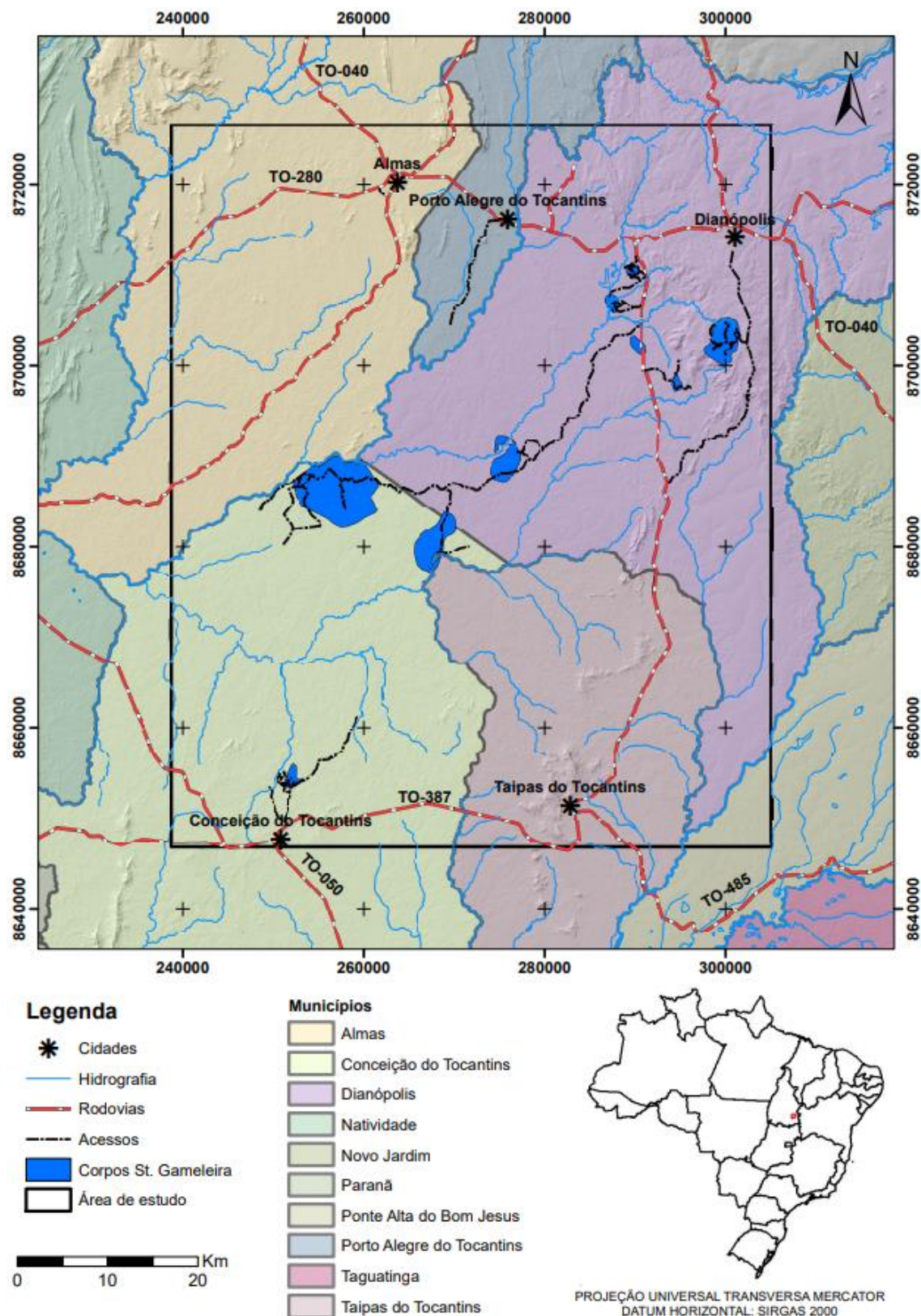


Figura 1: Mapa de localização indicando as principais rodovias e estradas de acesso, cidades, e redes drenagem inseridas na área de estudo. Em destaque em azul para os corpos já cartografados da Suíte Gameleira (Saboia et al., 2019; Abdalah et al., 2015). Modelo digital de elevação gerado pela imagem do Satélite ALOS sensor PALSAR indicando a topografia da área de estudo.

1.4 Estrutura da Dissertação

Esta dissertação foi dividida em 6 capítulos, onde:

O Capítulo 1 visa introduzir e contextualizar o tema da dissertação, bem como apresentar os objetivos gerais e específicos da pesquisa.

O Capítulo 2 contém de maneira ampla o contexto geológico no qual a área de estudo está inserida, contextualizado de maneira geotectônica e caracterizando as unidades que ocorrem na área.

O Capítulo 3 traz uma breve revisão sobre o aprendizado de máquinas (*Machine Learning*) e alguns conceitos que serão utilizados ao longo da dissertação.

O Capítulo 4 contém os materiais e métodos utilizados durante a pesquisa, trazendo informações sobre os métodos aerogeofísicos utilizados, a elaboração da base de dados e as características da implementação do *Random Forest*.

O Capítulo 5 apresenta o artigo elaborado durante a pesquisa, em que foram gerados mapas preditivos com ênfase na Suíte Gameleira e algumas formas de confirmações indiretas e diretas utilizadas.

O Capítulo 6 traz uma síntese dos resultados obtidos, e algumas considerações finais frente a pesquisa desenvolvida.

2. Geologia Regional

A Província Tocantins (de Almeida *et al.*, 1981) está localizada no centro do Brasil e corresponde a um sistema orogenético Neoproterozoico (650–570 Ma) que envolveu a convergência e colisão dos crátons: Amazônico a noroeste e São Francisco a leste. Segundo Fuck *et al.*, 2017 os movimentos de convergência e colisão tiveram fim no Neoproterozóico “*tardio*” consolidando a Província Tocantins como parte do processo de amalgamação da parte oeste do supercontinente *Gondwana*. Dois blocos continentais estáveis são relacionados ao ciclo orogênico Neoproterozóico, são eles o Bloco Parnaíba (Castro *et al.*, 2013), a nordeste e o Paranapanema (Mantovani & Brito Neves, 2009), a sudoeste, encobertos durante o Fanerozoico pelas bacias do Parnaíba e Paraná, respectivamente.

Três cinturões orogênicos Neoproterozóico são individualizados como principais integrantes da Província Tocantins: a) Faixa Araguaia, no limite leste do Cráton Amazônico; b) Faixa Paraguai, no limite sul do Cráton Amazônico, e a; c) Faixa Brasília, exposta ao longo da margem oeste do Cráton São Francisco (CSF) (Fuck *et al.*, 2014). A área de estudo do projeto é situada na porção norte da Faixa Brasília.

2.1 Faixa Brasília

A Faixa Brasília (FB) se entende por cerca de 1000 quilômetros com uma direção aproximadamente N-S (Fuck *et al.*, 2014). Ela é subdividida em Zona Cratônica, Zona Externa e Zona Interna, sendo que as duas últimas zonas são delimitadas pela Falha do Rio Maranhão, nessas zonas estão inseridos o Núcleo Metamórfico, Maciço do Goiás e o Arco Magmático do Goiás (Fuck *et al.*, 2005, 2014). A zona externa, localizada a leste da falha, é composta por espessas sequências de rochas sedimentares de margem passiva, metamorfizadas em baixo grau, e seu embasamento, caracterizado por terrenos granito-gnáissicos pertencentes ao bloco Crustal Cavalcante-Natividade (Brito Neves & Cordani, 1991; Fuck *et al.*, 1994, Pimentel *et al.*, 2011), sendo exposto devido à menor espessura das sequências metassedimentares. As rochas dessa porção da FB estão estruturadas em um sistema de empurrões e dobramentos com vergência em direção ao Cráton do São Francisco.

A FB possui uma evolução diacrônica que resultou na formação de dois segmentos com direções distintas: a Faixa Brasília Norte (FBN) de direção SW-NE, e a Faixa

Brasília Sul (FBS) com direção SE-NW. A separação entre os segmentos é estabelecida pela Sintaxe dos Pirineus, que marca a mudança das direções estruturais e configura a superposição de estruturas do segmento norte ao segmento sul (Araújo Filho, 2000).

A região de estudo está situada na porção setentrional da Zona Externa da Faixa Brasília no bloco Crustal Cavalcante-Natividade, interpretado como o embasamento siálico das rochas supracrustais paleoproterozoicas e neoproterozoicas da FBN. A compartimentação geotectônica da porção norte da Província Tocantins onde o bloco crustal foi subdividido em dois domínios: o Domínio Almas-Conceição do Tocantins e Domínio Cavalcante-Araías (Fuck *et al.*, 2014; Sousa *et al.*, 2016; Cordeiro & Oliveira, 2017) (Figura 2).

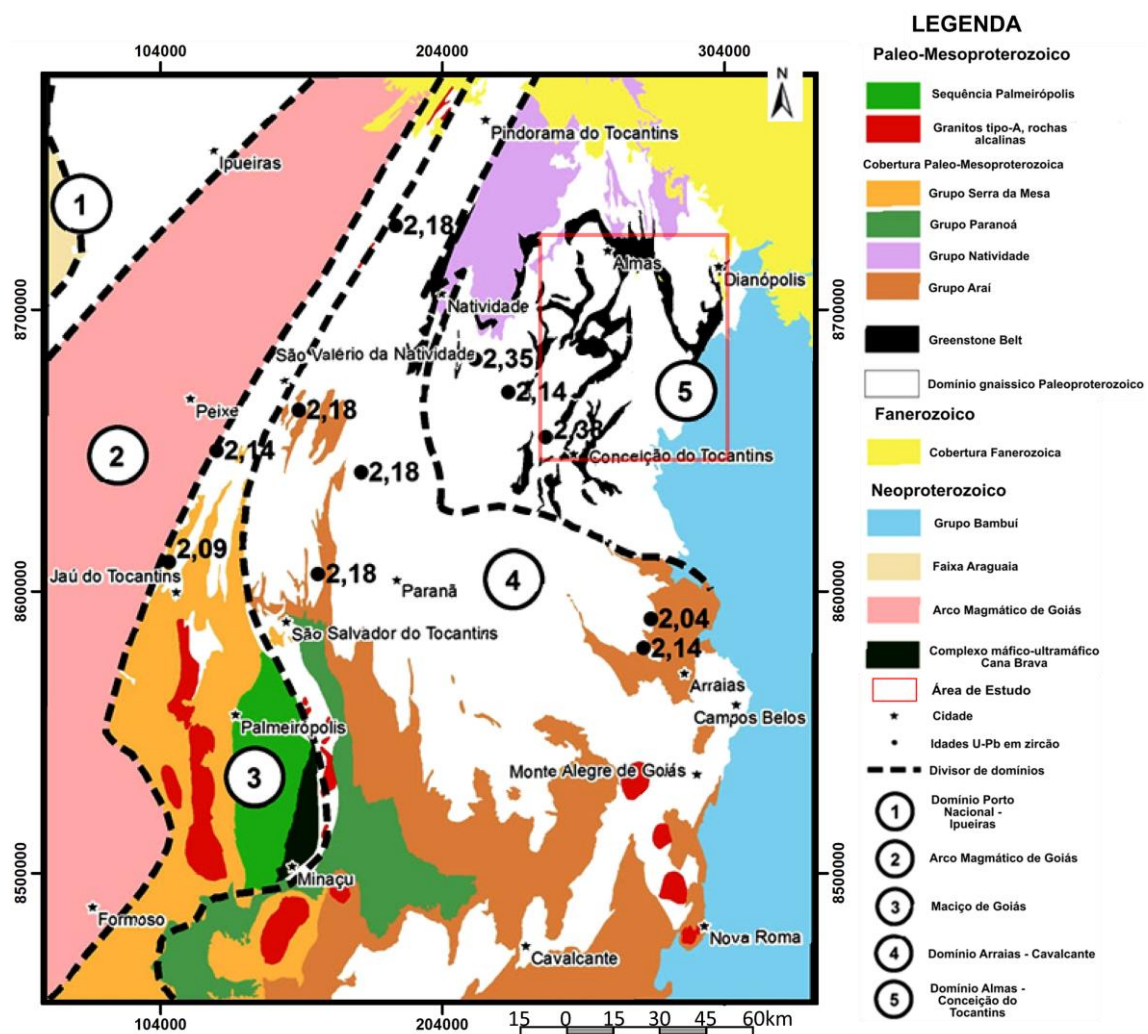


Figura 2: Mapa geológico esquemático do terreno Almas-Dianópolis com idades U-Pb em evidência e delimitando os diferentes domínios tectônicos (Adaptado de Fuck *et al.*, 2014). O retângulo em destaque vermelho indica a área de estudo inserida no domínio 5, a unidade nomeada Greenstone Belt no mapa corresponde ao Grupo Riachão do Ouro e a Suíte Gameleira.

2.2. Bloco Crustal Cavalcante-Natividade (BCCN)

Dois grandes domínios são reconhecidos como formadores e constituintes deste bloco crustal. Na parte nordeste do BCCN é localizado o Domínio Almas-Conceição do Tocantins (DACT), caracterizado por uma ampla extensão de granitos-gnaisses associados a sequências metassedimentares, e um *trend* regional preferencial N-S. Essas sequências estão dispostas em “faixas” envolvendo os domos graníticos-gnáissicos principalmente a norte na região próxima a Almas-Dianópolis (Fuck, 2014; Costa *et al.*, 1976; Costa, 1985; Cruz & Kuyumjian, 1998, 1999; Borges *et al.*, 1998; Kuyumjian *et al.*, 2012). Essa associação foi originalmente interpretada como um Terreno Arqueano Granito- *Greenstone* (Fuck, 2014; Costa *et al.*, 1976; Padilha, 1984). Idades radiométricas pelo método Rb-Sr resultaram em idades arqueanas entre 2,5 a 2,9 Ga em rochas coletadas próximas à cidade de Dianópolis, uma rocha vulcânica félsica pertencente ao conjunto de rochas das “faixas” supracrustais próximas a cidade de Natividade, foi datada pelo método U-Pb, obtendo uma idade de 2206 ± 13 Ma (Dardenne *et al.*, 2009). O DACT está em contato com o Domínio Cavalcante-Arraias por contato tectônico por meio da Zona de Cisalhamento Cruz das Almas (zona de cisalhamento transcorrente dextral), que possui um mergulho moderado a alto para W-NW (Praxedes, 2015).

A porção remanescente do BCCN, denominado Domínio Cavalcante-Arraias (Fuck *et al.*, 2014) estende-se desde Cavalcante, a sul, até Arraias, Paranã, a oeste, e além de Natividade até Pindorama do Tocantins, a norte. Esse domínio tem a ocorrência predominante de granitoides peraluminosos. Por mais que ocorram rochas peraluminosas em ambos os domínios, a relação entre as rochas de ambos os domínios ainda é incerta, bem como os limites entre os domínios, uma vez que granitos peraluminosos intrudindo ortognaisses metaluminosos próximos à cidade de Natividade foram mapeados (Oliveira, 2012). O presente trabalho tem sua área majoritariamente no Domínio Almas-Conceição do Tocantins, o foco, do mesmo, será dado as unidades que compõem o mesmo.

2.2.1 Embasamento Paleoproterozoico - Domínio Almas-Conceição do Tocantins (DACT)

Autores colocam o DACT como a porção mais a norte e menos compreendida do Maciço do Goiás (Fuck *et al.*, 2014; Sousa *et al.*, 2016; Cordeiro & Oliveira, 2017) relacionando-o a um ambiente de subducção no início do Paleoproterozoico (Cruz e Kuyumjian, 1998; Kuyumjian *et al.*, 2012).

O embasamento arqueano/Paleoproterozoico é representado no Domínio Almas-Conceição do Tocantins por rochas ortognaissicas, granitoides, rochas metavulcanossedimentares dispostas em faixas em forma de ‘Y’, de direção NS com ramificações para NE e NW (Kuyumjian *et al.*, 2012). Rochas neoproterozoicas pertencentes aos grupos Bambui, bem como rochas paleo à mesoproterozoicas do grupo Natividade ocorrem na região do DACT de maneira discordante as rochas do embasamento (Pimentel *et al.*, 2004; Sabóia, 2009). Unidades são sobrepostas localmente por coberturas cenozoicas detrítico-lateríticas ferruginosas, depósitos aluvionares e intrudidas por diques cretáceos (Saboia, 2014, 2019, 2021).

A área de estudo encontra-se predominantemente dentro do DACT, onde somente a sudeste faz contato com rochas do Grupo Bambui, pertencente ao domínio Cratônico, as demais unidades se inserem no contexto do DACT (Figura 3). Nos próximos tópicos serão descritas as unidades encontradas na área de estudo.

2.2.1.1 Grupo Riachão do Ouro

O Grupo Riachão do Ouro (GRO) foi primeiramente descrito como rochas vulcanossedimentares supracrustais, parte do Terreno granito-*greenstone* do Tocantins (Costa, 1984). O grupo foi compartimentado em duas unidades, uma unidade basal composta por rochas vulcânicas metamáficas e metaultramáficas toleíticas e komatiíticas denominada Formação Córrego do Paiol, e a Formação Morro do Carneiro que corresponde a unidade de topo, composta por uma sequência de rochas metassedimentares e sedimentares químicas. Ambas formações estão em um intervalo entre as fácies metamórficas xisto verde e anfibolito e zonas de cisalhamento associadas a unidade apresentam importantes ocorrências auríferas (Cruz & Kuyumjian, 1998; Kuyumjian *et al.*, 2005, 2012).

O GRO corresponde a unidade mais antiga do DACT, com idade mínima de 2,47 -2,45 Ga com base na datação de granitoides da unidade Ribeirão das Areias que contem xenólitos de rochas do GRO (Cruz *et al.*, 2003; Martins-Ferreira *et al.*, 2017). Idades arqueanas paleoproterozoicas a partir de isótopos de Nd indicam uma idade T_{DM} de 2,51 Ga para metabasaltos pertencentes a Formação Córrego do Paiol (Saboia & Meneghini, 2019). A idade mínima para o Grupo Riachão do Ouro é baseada em cima da datação de um xenólito da unidade dentro de granitoides pertencentes ao Complexo Ribeirão das

Geologic Unities

- Cover
- Cenozoic Alluvion
- Cretaceous dikes

Neoproterozoic

- Bambui Group

Mesoproterozoic

- Natividade Group

Almas - Conceição do Tocantins Domain

Rhyacian Unities (>2.3 - 2.05 Ga)

- Gameleira Suite
- Peraluminous Suite
- Peraluminous - Tonalitic Suite
- Serra do Boqueirão Suite
- Granodioritic-Tonalitic Suite
- Quartz-Dioritic Suite
- Monzogranitic Unit

Siderian Unities (2.47 - 2.3 Ga)

- Ribeirão Itaboca Suite RIS (2.3 Ga)
- Ribeirão das Areias Complex RAC (2.47 - 2.45 Ga)
- Ribeirão das Areias Complex (RAC) - IB

Riachão do Ouro Group

- Morro do Carneiro
- Córrego do Paíol Formation

Tectonic Structures

- contractional fault
- extensional fault
- fault or approximate shear zone
- strike-slip fault
- dextral strike slip shear zone
- sinistral strike slip shear zone
- compressional shear zone
- non-defined shear zone
- strike slip shear zone
- transpressional shear zone
- fault or non-defined shear zone
- approximate strike slip shear zone

Map Labels: Arraias, Pôrto Alegre do Tocantins, Chacaraçu, Rio das Aranhas, Rio Manoel Alves, Rio Camargo, Rio Ilhéus, Rio Jiboca, Rio Augusto, Taipas do Tocantins, Genorção do Tocantins.

Scale: 0 5 10 km

Coordinates: X-axis: 240000, 260000, 280000, 300000; Y-axis: 8720000, 8700000, 8680000, 8660000.

A Formação Córrego do Paiol (FCP) é composta predominantemente de metabasaltos com raras ocorrências de rochas ultramáficas, são localmente encontradas estruturas do tipo “*pillow lava*” preservadas, indicando a presença de um vulcanismo máfico subaquoso (Cruz & Kuyumijan, 1999; Kuyumijan *et al.*, 2012). Nas vulcânicas máficas individualizou-se dois tipos de magmas fonte, os magmas toleíticos ricos em Fe e subordinados basaltos komatiitos ricos em Mg. As características geoquímicas das rochas indicam um ambiente tectônico referente a arcos magmáticos recentes (Abdallah, 2015).

Já a Formação Morro do Carneiro (FMC) é situada no topo da sequência metavulcanossedimentar do Grupo Riachão do Ouro e é formada por: metassedimentos terrígenos e químicos, composta por filito sericítico carbonoso, quartzito (podendo ocorrer localmente por vezes com magnetita ou com turmalina); metassilexito, xisto carbonoso, formação ferrífera bandada, metaconglomerado e, subordinadamente, rochas metavulcânicas félsica (muscovita-plagioclásio-quartzo xisto), intermediária (clorita-anfibólio xisto feldspático) e máfica (anfibólio xisto e anfibolito) (Cruz & Kuyumjian, 1998; Cruz & Kuyumjian, 1999; Kuyumjian *et al.*, 2012).

2.2.1.2 Complexo Granito-Gnáissico – Complexo Almas- Cavalcante

Este terreno é composto por associações de granitoides, onde Correia Filho & Sá (1980) subdividiu a ocorrência dessas rochas no DACT em duas associações: Conjunto Granítico-Migmatítico e Conjunto Migmatítico-Anfibolítico. Este conjunto de rochas já foi chamado por outros autores como Complexo Goiano (Costa, 1985) e Complexo Alto Paranã (Borges, 1998). Neste trabalho a nomenclatura utilizada será Complexo Almas-Cavalcante, para representar o conjunto de intrusões, o mesmo utilizado pelas Folhas Arraias (Abdalah & Meneghiini, 2017) e a Folha Dianópolis (Saboia & Meneghini, 2019),

Três tipos de rochas intrusivas no Grupo Riachão do Ouro que apresentam baixo potássio foram descritos por Cruz e Kuyumjian (1998) e Cruz *et al.*, (2003), são eles: Suíte 1 com caráter metaluminoso, apresenta pouco alumínio, rica em anfibólio (hornblenda), composta predominantemente de tonalitos. Possui idades U-Pb a partir de zircão de 2,2 Ga e dados de isótopos de Nd com TDM = 2,54 a 2,45 Ga e $\epsilon\text{Nd}(t) = -1,37$ a $-0,15$ (Saboia, 2021). Os autores consideram que a Suíte foi formada a partir da fusão parcial de uma fonte ultramáfica residual em um ambiente de cunha mantélica sobre uma zona de subducção; Suíte 2 é peraluminosa e apresenta baixa abundância de minerais máficos, tendo a biotita como mineral máfico dominante, composta por tonalitos, trondhjemitos, granodioritos e subordinadamente monzogranitos. Possui idades U-Pb a partir de zircão de 2,2 Ga e dados de isótopos de Nd com TDM = 2,76 a 2,53 Ga e $\epsilon\text{Nd}(t) = -4,98$ a $-0,88$; Complexo Ribeirão das Areias é semelhante a Suíte 2 se diferenciando por suas idades U-Pb a partir de Titanita de 2,45 Ga e $\epsilon\text{Nd}(t) = +0,58$ (Saboia, 2021). Os mesmos autores colocam a origem desses momentos magmáticos como decorrente da

fusão parcial de metabasaltos de crostas oceânicas subductadas ou de novas *underplated* crostas basálticas.

No presente trabalho a subdivisão de unidades para este terreno utilizada foi a proposta por Saboia (2021), em que o autor subdivide o terreno em unidades tendo como base estudos geocronológicos. As unidades de idade Sideriana correspondem ao Complexo Ribeirão das Areias (2,47 Ga) e a Suíte Itaboca (2,3 Ga) proposta por Saboia, 2020a, 2020b. Já as unidades de idades Riagianas que ocorrem dentro do DACT são: Unidade Monzogranítica (2,29 Ga), Suíte Quartzo – Diorítica (2,28 Ga), Suíte Granodiorítica a Tonalítica (2,26 Ga), Suíte Serra do Boqueirão (2,22 Ga) e Suíte Peraluminosa (2,2 Ga) (Cruz *et al.*, (2003; e referências nele contidas); Abdallah & Meneghini, 2017; Sousa *et al.*, 2016; Saboia, 2021).

Uma breve descrição das unidades do Complexo Almas-Cavalcante utilizadas neste trabalho será realizada a seguir, a mesma teve como base o trabalho de Saboia (2021):

i) Complexo Ribeirão das Areias (RAC): metatonalitos mesocráticos e leucocráticos e metatrandhjemitos com ocorrências subordinada de granodioritos. Ocorrem como batólitos com formato sigmoidal orientados NNW-SSE a NNE-SSW, no núcleo desses batólitos estruturas primárias encontram-se preservadas. Essa unidade possui uma subunidade mapeada como Ribeirão das Areias, corpos intrusivos (RAC-ib), descrita como um metagranodiorito de granulação média a grosseira e intrusões monzograníticas.

ii) Suíte Ribeirão Itaboca (RIS): é a unidade com maior área de ocorrência no projeto, é exposta nas porções mais arrasadas, as rochas que compõem essa unidade são biotitas metatonalitos e gnaisses tonalíticos. Em trabalhos anteriores essa unidade era inserida na associação anfilítica-migmática, como complexo granítico gnáissico não subdividido ou como parte dos granitoides da Suíte 1 e 2 (Correia Filho and Sá, 1980; Cruz *et al.*, 2003). Essa unidade é caracterizada por apresentar feições de migmatítico gnáissico.

iii) Unidade Monzogranítica (UnMonz): descrita como um conjunto de corpos graníticos semicirculares a elípticos, composta principalmente por metagranitos de granulação média a grossa, inequigranular a porfirítica. O corpo granítico Pau Ramalhudo (Correia Filho e Sá, 1980) é a intrusão mais característica desta unidade, orientado NE-SW, faz contato direto com o corpo mais expressivo da Suíte Gameleira.

iv) Suíte Quartzo-Diorítica (SQD): composta de dois grandes corpos e algumas intrusões de menor porte de quartzo-metadiorito, metagabros e subordinadamente metatonalitos (Abdallah and Meneghini, 2017).

v) Suíte Granodiorítica a Tonalítica (SGT): esta classe é composta por parte da Suíte 1 descrita por Cruz *et al.*, (2003), aonde parte da Suíte Manoel Alves e a Suíte Rio do Moleque definida por Corrêa *et al.*, (2015) e a Suíte Conceição do Tocantins proposta por Sousa *et al.*, (2016) (Saboia, 2021). Os litotipos predominantes dessa unidade são biotitas metagranodioritos e metatonalitos, com ocorrência secundária e hornblenda-biotita metatonalitos, biotita metamonzogranitos e quartzo metadiorito.

vi) Suíte Serra do Boqueirão (SSB): é caracterizada por um hornblenda metatonalito de granulação média a grossa, mesocrático, com texturas inequigranulares e foliação. Sua área tipo corresponde ao Plúton Corrente descrito por Correia Filho e Sá, (1980), próximo à cidade de Dianópolis.

vi) Suíte Peraluminosa (SP): a unidade está contida nos granitoides da Suíte 2 conforme Cruz *et al.*, 2003, e classificada como uma unidade formada por rochas peraluminosa podendo ser metagranodioríticas ou metatonalíticas. Saboia (2021) individualizou duas litofácies: (i) muscovita-biotita metagranodiorito com ocorrência subordinada de metatonalitos e metamonzogranitos, e (ii) *stocks* de metatonalito leucocrático com menos de 10% de biotita (SPT).

2.2.1.3 Suíte Gameleira (SGm)

Os primeiros trabalhos que remetem a unidade datam de 1976 e dizem respeito a Suíte Máfico-Ultramáfica Caraíbas, que correspondeu ao agrupamento dos maciços de Barra do Gameleira, Marta-Tamboril e Cerqueira Santaninha (Costa, 1976). Correia Filho & Sá (1980) individualizaram a unidade em Maciço Básico-Ultrabásico Barra do Gameleira subdividindo em duas categorias a unidade: Gabro-Peridotítica Marginal e Gabro-Granulítica Central. A denominação Intrusivas Máfico-Ultramáficas Tipo Gameleira foi realizada por Danni & Texeira (1981), e é utilizada na Folha Dianópolis (Saboia & Meneghini, 2019), já na Folha Arraias (Abdallah & Meneghini, 2017) utilizou-se o nome Suíte Máfico-Ultramáfica Gameleira para a cartográfica. Saboia (2021) utiliza o termo Suíte Gameleira, esta é a nomenclatura utilizada neste projeto.

Os corpos desta suíte correspondem a intrusões acamadadas, com forma semicircular a corpos alongados. Estes corpos foram submetidos, localmente, a um

metamorfismo em fácies anfibolito a granulito (Correia Filho e Sá, 1980, CPRM, 2000) e apresentam-se deformados, preservando em certos domínios estruturas ígneas primárias. Encontra-se em contato com as rochas do Complexo Almas-Cavalcante e do Grupo Riachão do Ouro de forma intrusiva ou por contato tectônico. A área tipo desta suíte, maior intrusão mapeada, foi compartimentada em domínios predominantemente máficos com o posicionamento de rochas ultramáficas relacionadas a zonas de cisalhamento (CPRM, 2000).

A Intrusão Barra do Gameleira (IBG) é referenciada neste trabalho como a maior intrusão da Suíte Gameleira e possui um formato semi-arredondado, com extensão aproximada de 8 quilômetros (N-S) x 10 quilômetros (E-W), apresentando estruturas e texturas ígneas preservadas (Figura 4). Na porção central da intrusão ocorre uma zona de cisalhamento local marcada por estruturas e texturas deformacionais e metamórficas, com uma foliação de atitude aproximada N20-25W/50-60SW. As rochas da IBG são caracterizadas por um forte retrometamorfismo em fácies xisto-verde, evidenciado por substituição de minerais primários (CPRM, 2000)

As rochas máficas, predominantes, são principalmente gabros, leucogabros, troctolitos, olivina-gabronoritos, olivina-gabros coroníticos e noritos. Localmente são observados gabros, microgabros e dioritos, com paragênese típica da fácies granulito (Correia Filho e Sá, 1980, CPRM, 2000). As rochas ultramáficas ocorrem subordinadamente em forma de lentes ou alternadas com as rochas gabróicas, seus principais representantes são dunitos, piroxenitos feldspáticos, peridotitos serpentinizados e serpentinitos. De maneira geral, estas rochas levam o prefixo ‘meta’ por estarem em um contexto de metamorfismo de fácies xisto-verde a anfibolito.

As rochas máficas da Suíte Gameleira possuem abundância de cristais de zircão, interpretados como contaminação crustal, apresentando uma idade $^{207}\text{Pb}/^{206}\text{Pb}$ entre 2,48 e 2,30 Ga (Saboia, 2021).

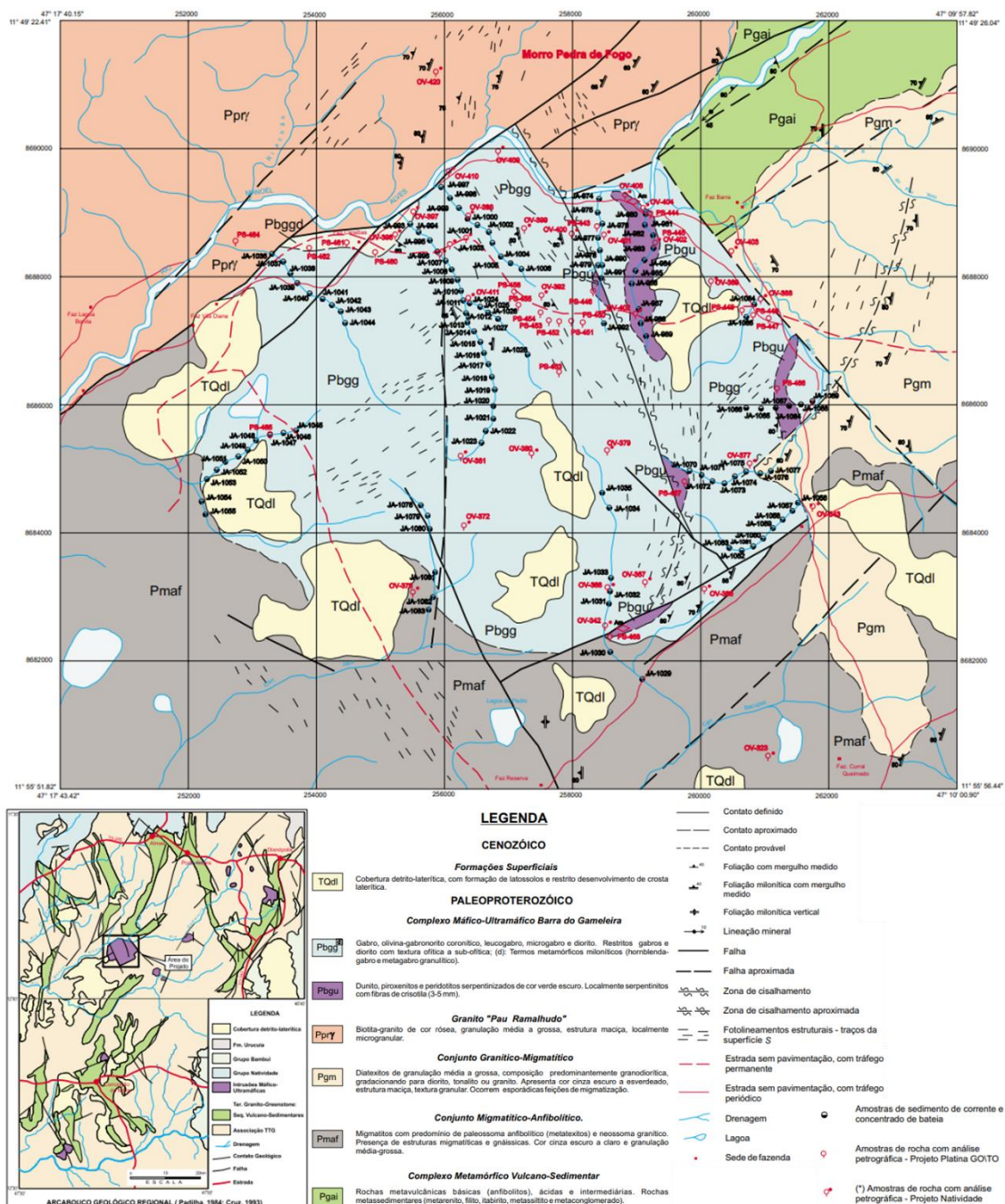


Figura 4: Resultados do Projeto: “Geologia e resultados prospectivos da Área Barra do Gameleira – Tocantins” (CPRM, 2000). A - Mapa geológico da intrusão Barra do Gameleira, predominância de rochas máficas e subordinamente ultramáficas controlados por estruturas. *Link para download:* <<https://rigeo.sgb.gov.br/bitstream/doc/1610/1/Geologia%20e%20resultados%20prospectivos%20da%20%20c3%a1rea%20Barra%20da%20Gameleira%20-%20Tocantins.pdf>>

2.2.1.4 Grupo Natividade (GrNat)

Os principais trabalhos realizados sobre o conteúdo litológico e a extensão do Grupo Natividade são de Portela *et al.*, (1985), Saboia *et al.*, (2019), Dardenne & Saboia (2007), Gorayeb *et al.*, (1988), Hasui *et al.*, (1990), Oliveira *et al.*, (2012) e Praxedes

(2015). A correlação com o Grupo Arai e consequentemente com o Grupo Serra da Mesa e Supergrupo Espinhaço, situa o Grupo Natividade no evento de rifteamento da Tafrogênese Estateriana entre 1,8 e 1,6 Ga (Brito Neves, 2002; Bizzi *et al.*, 2003; Marques, 2009; Pimentel *et al.*, 1991).

Na área de estudo as ocorrências desse grupo estão na porção norte da área e todas as unidades, a seguir, foram compiladas na classe Grupo Natividade, as unidades que ocorrem na área conforme a Folha Dianópolis (Saboia & Meneghini, 2019) são: (i) *PP4na1* – É representada por fácies de quartzitos e metarenitos com intercalações de metaconglomerados suportados por matriz, além de metadolomitos restritos. (ii) *PP4na2* - metapelitos com fácies de contribuição psamíticas subordinadas; (iii) *PP4na2cc* - é composta por carbonatos impuros micáceos (calcixistos) e metacalcários de cor cinza; (iv) *PP4na2qt* - é constituída por metarenitos e quartzitos silicificados estratificados em bancos decimétricos a métricos, além de metaconglomerados subordinados; (v) *PP4na3* - predomina metapelitos com intercalações de metapsamitos compondo metarritimitos micáceos de cor cinza-esverdeada.

2.2.1.5 Grupo Bambuí (GrBam)

O Grupo Bambuí foi formado por ciclos de sedimentação carbonático pelito-arenoso, onde seus depósitos se correlacionam a uma extensa transgressão marinha avança sobre o Cráton São Francisco, no Neoproterozoico (Alkmim & Martins-Neto, 2012). A divisão estratigráfica do grupo proposta por Costa & Branco (1961) foi adaptada por Dardenne (1978), que individualiza da base para o topo: Sete Lagoas, Serra de Santa Helena, Lagoa do Jacaré, Serra da Saudade e a Formação Três Marias. Idades em torno de 506 Ma e 625 Ma foram encontradas somente na Formação Sete Lagoas, se forma que esses dados indicam uma idade máxima para sedimentação do Grupo de 560 Ma (Paula-Santos *et al.*, 2015).

Na área de estudo ocorrem duas unidades pertencentes ao Grupo Bambuí, são elas: (i) Formação Sete Lagoas (NP2sld) – composta por argilitos laminados e siltitos finos na base, sotopostos por metacalcarenitos, metadolarenitos e metadolomitos micríticos e subordinadamente. A litofácies carbonática predominante corresponde aos calcários dolomíticos e dolomitos, que formam uma faixa comumente sobreposta aos pelitos síltico-argilosos e assentada sobre o embasamento granítico; (ii) Formação Serra de Santa Helena (NP2shp) – predominam sedimentos clásticos que representam a sedimentação pelito-carbonática do Grupo Bambuí. Constituída por ardósias

estratificadas e de cor acinzentada e metapelitos de composição síltico-argilosa, exibindo clivagem ardosiana. Metarritimitos são formados por intercalações de metassiltitos e metargilitos.

3. Aprendizado de Máquinas

A inteligência artificial é o campo que utiliza algoritmos computacionais para elaboração de máquinas inspiradas na inteligência de interpretação humana (Angelov *et al.*, 2021). O conceito de inteligência artificial abrange vários métodos a depender do nível de complexidade e capacidade da automação, se organizando de maneira hierárquica por ordem de sua complexidade (Figura 5). Dentre eles há um método que faz com que o computador simule um processo de aprendizagem a partir de um conjunto de dados fornecidos, contendo “rótulos” conhecidos, que possibilitam a geração de resultados preditivos ou descritivos, este método é chamado de aprendizagem supervisionada, e será o foco deste trabalho. Um outro método seria a aprendizagem não supervisionada, nesse não são cedidos dados para treinamento e é realizado, por exemplo, um agrupamento estatístico de todos os dados. Ambos os métodos fazem parte do que chamamos de Aprendizado de Máquinas (AM) ou *Machine Learning* (ML).

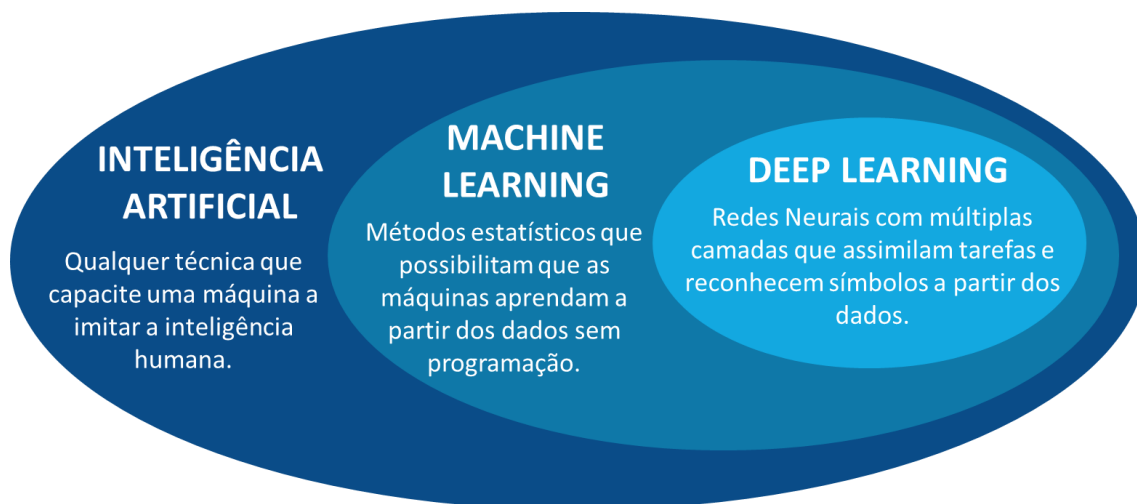


Figura 5: Representação hierárquica dos diferentes conceitos da inteligência artificial. Fonte: <<https://opencadd.com.br/machine-learning-ou-deep-learning/>>, acessado em 17/04/2023.

Uma das primeiras definições de AM foi realizada por Arthur Lee Samuel em 1959, responsável pelo programa de jogo de damas, em que o programa conseguiu aprender o jogo a ponto de apresentar um exímio desempenho. O mesmo, definiu AM como o "campo de estudo que dá aos computadores a habilidade de aprender sem serem explicitamente programados". Tom Mitchell em 1983 apresentou uma definição mais moderna e formal para o conceito de AM: “Diz-se que um programa de computador aprende com a experiência e com relação a alguma classe de tarefas T e medida de

desempenho P, se seu desempenho em tarefas em T, conforme medido por P, melhora com a experiência E.” (Michalski *et al.*, 1983).

A depender do conjunto de dados a serem utilizados como entrada e saída, e o tipo de problema que se visa resolver, o tipo de algoritmo de AM deve ser escolhido. Existem três categorias principais de abordagem para o processamento de dados no Aprendizado de Máquinas: (a) aprendizado supervisionado, aplicações nas quais os dados de treinamento compreendem exemplos sobre os vetores de entrada com informações sobre seus vetores-alvo (rótulos). Casos em que se planeja atribuir a cada vetor de entrada um número discreto de categorias, são chamados de problemas de classificação. Se a saída consiste em uma ou mais variáveis contínuas, então o problema é de regressão; (b) aprendizado não-supervisionado, nestes tipos de problema os dados de treinamento consistem em um conjunto de entradas sem nenhuma informação sobre o vetor-alvo (rotulo). O algoritmo então observa e identifica padrões de distribuição nos dados de entrada e os agrupa com base em seus atributos (*clustering* e redução de dimensionalidade); (c) aprendizado por reforço (Sutton e Barton, 1998): aqui o algoritmo não recebe exemplos de saídas corretas, mas buscam encontrá-las a partir de um processo de tentativa e erro, onde o algoritmo é treinado a partir de recompensas e punições a depender da ação a ser tomada. (Russel & Norvig, 2021).

Conforme a área de estudo, os modelos apresentam-se de formas distintas, simples e organizados ou complexos e mal organizados, podendo ter poucas variáveis ou muitas variáveis a ponto de suas influências não serem completamente quantificadas (Molugaram, *et al.*, 2017). Nas ciências da terra os sistemas trabalhados, quase sempre, correspondem a sistemas não lineares, de forma que apresentam inúmeras variáveis. Dessa maneira o grupo de “dados de treino” abrangente é construído buscando cobrir o máximo possível de parâmetros do sistema (Lary, 2010). Nesse cenário, o uso do AM é ideal para resolver os problemas no qual a teoria do conhecimento ainda é incompleta, mas para o qual temos um significativo número de observações e outros dados.

Dito isso, o foco deste estudo é em cima do aprendizado supervisionado, em que a premissa básica é que os dados de treino contenham rótulos e informações da classe inferida (Kotsiantis, 2007). A arquitetura dos algoritmos de aprendizado de máquina e as distribuições estatísticas dos dados observados, por vezes balizam o treinamento dos modelos de classificação, que age minimizando a função perda (erro). Dessa maneira, os modelos de classificação treinados são atrelados as variáveis de entrada semelhantes, para

predizer a que tipo de categoria tal entrada pertence, com base nos dados de treino (Cracknell, 2013).

Os algoritmos de aprendizado de máquinas (AAM) vem sendo aplicados utilizados em diversas áreas das geociências, visando resolver e propor novas hipóteses para uma variedade de problemas, como, por exemplo: proveniência de solos (Aitkenhead, 2014; Ma, 2021); mapas de prospectividade mineral (MPM) (Zhang, 2018; Parsa, 2023); análise mineral usando dados de furos de sonda ou amostras (Jung & Choi, 2021), dentre outras áreas das geociências e suas esferas do conhecimento. O uso de AM para a realização de predições litológicas vem sendo uma das grandes frentes no uso de AAM nas geociências (Cracknell & Reading, 2013, 2014a, 2014b; Harris, 2012, 2015; Kuhn, 2018, Radford, 2018; Costa, 2019, 2020; Kuhn, 2018, 2019), e tem sido uma abordagem amplamente aceita e complementar ao trabalho de campo clássico, uma vez que apresenta um bom custo-benefício especialmente em áreas de difícil acesso. Neste contexto, uma forma de maximizar a cobertura do mapeamento é utilizando dados de sensoriamento remoto (Radar), dados aerogeofísicos e dados geoquímicos (Harris, 2015).

Na literatura encontram-se trabalhos com o uso de dados aerogeofísicos e dados de sensoriamento remoto, cujo objetivo é avaliar qual o melhor algoritmo para geração de um mapa preditivo litológico (Cracknell & Reading, 2014a, e Costa *et al.*, 2019), de forma que cinco algoritmos de aprendizado de máquina foram testados, sendo esses: *Naive Bayes* (NB) – algoritmo de aprendizado estatístico, *k-Nearest Neighbors* (kNN) – aprendizado baseado em instâncias, *Random Forests* (RF) – aprendizado baseado em lógica, *Support Vector Machines* (SVM), e *Artificial Neural Networks* (ANN).

A partir destes trabalhos, concluiu-se que o algoritmo *Random Forest* apresenta a melhor desempenho quando comparado aos demais algoritmos testados, por apresentar uma melhor relação de métricas de predição do modelo, bem como ser um algoritmo mais robusto para *outliers* e amostras não balanceadas (Breiman, 2001). Cracknell & Reading, 2014a, utilizam tanto dados aerogeofísicos, como Modelos Digital de elevação e dados decorrentes de sensoriamento remoto, enquanto Costa *et al.*, 2019, não utilizam dados decorrentes de sensoriamento remoto. Com base nestes trabalhos e em outros (Harris, 2012, 2015; Kuhn, 2018, Radford, 2018; Costa, 2020; Kuhn, 2018, 2019), o presente estudo terá o foco no uso do algoritmo *Random Forest*.

3.1 Random Forest

É um algoritmo de aprendizado de máquinas supervisionado do tipo *ensemble* (conjunto), onde é gerada uma combinação de árvores de classificação e regressão no qual a cada uma é fornecida um vetor amostrado independentemente, de forma aleatória, e apresenta a mesma distribuição (Breiman, 2001).

Um dos paradigmas que envolve o aprendizado do tipo *ensemble* é o *bagging* (ou *bootstrapping aggregating*), o algoritmo de aprendizado *Random Forest* baseia-se nesse método para reduzir a variância da classificação. O método *Bagging* atua no algoritmo criando múltiplas cópias das amostras do banco de dados de treino (cada cópia realizada é um pouco diferente da outra) (Burkov, 2019) e, a partir disso, as aplica em múltiplas árvores semelhantes às árvores de regressão e classificação (CART, *Classification and Regression Tree*) (Breiman *et al.*, 1984, Breiman, 1996). Cada árvore é treinada com os dados remostrados a partir dos dados originais (método *bootstrapping*) para encontrar um subconjunto de variáveis selecionadas aleatoriamente para a subdivisão de cada “nó” (Figura 6).

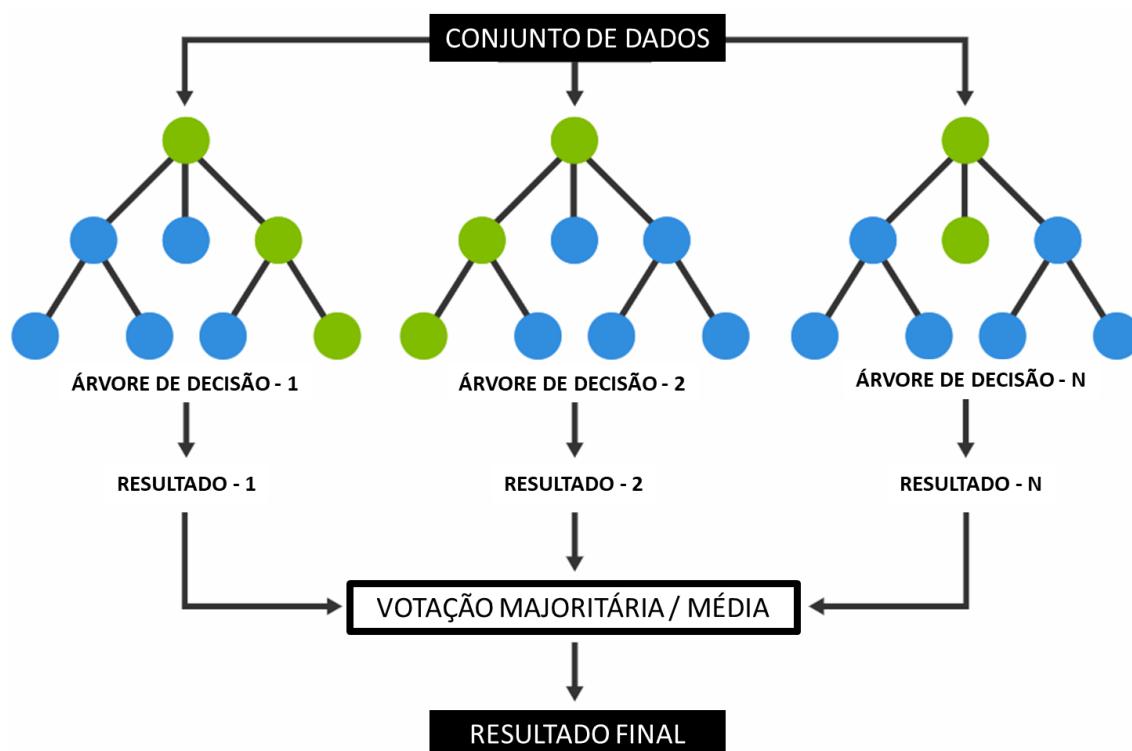


Figura 6: Processo de decisão de classe pelo algoritmo Random Forest. Fonte: <<https://www.tibco.com/pt-br/reference-center/what-is-a-random-forest>>, acessado em 17/04/2023.

Esse método é empregado para evitar a correlação entre árvores: se um ou mais parâmetros são excelentes para a predição da classe alvo, esses parâmetros serão

selecionados e colocados em decisão por várias árvores (Burkov, 2019), de maneira que o voto majoritário resultante das diversas árvores de decisão aleatórias seja utilizado para gerar a predição da classe (Cracknell & Reading, 2013). O fator de correlação entre múltiplas árvores de decisão aleatórias é o responsável pelo nome do algoritmo “Florestas Aleatórias” (*Random Forest*).

O algoritmo *Random Forest* gera resultados com precisão para um número limitado de amostras de treinamento, além disso, é estável as variações dos parâmetros do modelo de classificação e insensível para dados “irrelevantes”. O algoritmo também fornece uma boa indicação de ambiguidades e erros de classificação através das incertezas de classificação (Cracknell & Reading, 2013), dando espaço para muitas vezes discutir o modelo existente frente a novidades e inconsistências do modelo predito.

Para uma melhor compreensão do algoritmo o conhecimento acerca de alguns conceitos e métricas de avaliação é fundamental.

3.2 Conceitos de Aprendizado de Máquinas

Ao realizar a etapa de treino do modelo, espera-se que este classifique corretamente os dados que não fazem parte do conjunto treino, apresentando boa capacidade de generalização. A maneira para se realizar essa análise de aderência do modelo é através das métricas resultantes do treinamento do modelo e da predição.

Para a confecção do modelo subdivide-se o conjunto de dados em três: **i) Conjunto de treinamento:** é a parte dos dados utilizados para treinar o modelo de aprendizado de máquina. Ele consiste em dados rotulados (entradas e saídas conhecidas); **ii) Conjunto de teste:** Corresponde aos dados que não foram utilizados no conjunto de treinamento. Os dados deste conjunto são utilizados para a realização de previsões geradas através do modelo e compará-las com os rótulos originais para avaliar o desempenho do modelo; **iii) Conjunto validação:** é uma parte adicional dos dados que é utilizada durante a etapa de treinamento, mas com a intenção de ajustar os hiperparâmetros do modelo e avaliar o seu desempenho em diferentes configurações.

3.2.1 Overfitting e underfitting

Para avaliar se há ou não o processo de underfitting, ou *overfitting* (Figura 7), utiliza-se o *dataset* previamente separado em dados teste. O *underfitting*, ou alto *bias*, é relacionado a modelos que realizam muitos erros ao ter seu desempenho avaliado nos

dados que foi treinado. Sendo assim, o *underfitting* acontece quando o modelo é incapaz de prever bem os rótulos dos dados nos quais foi treinado. Alguns dos possíveis motivos para isso ocorrer são: o modelo é muito simples para o *dataset*; o conjunto de dados não são tão informativos a ponto de gerar boas previsões. Uma das formas de contornar essa situação é tentar usar um modelo mais complexo ou melhorar a qualidade dos seus dados, tornando-os mais informativos (Burkov, 2019).

O problema inverso seria o *overfitting*, problema de alta variância, o modelo comete quase erro nenhum em relação aos dados de treino, de maneira que se os dados de treino fossem randomizados de maneira distinta, o aprendizado resultaria em um resultado significativamente diferente. Este é o motivo no qual modelos que apresentam *overfitting* possuem desempenho fraco frente os dados de teste. Essa situação pode ocorrer caso seu modelo seja muito complexo para seu *dataset* ou seu *dataset* apresentar muitos *features* para um número pequeno de dados de treino. Para contornar esta situação pode-se tentar usar modelos mais simples, aumentar o volume dos dados de treino, ou reduzir a dimensionalidade dos exemplos no *dataset* (Burkov, 2019).

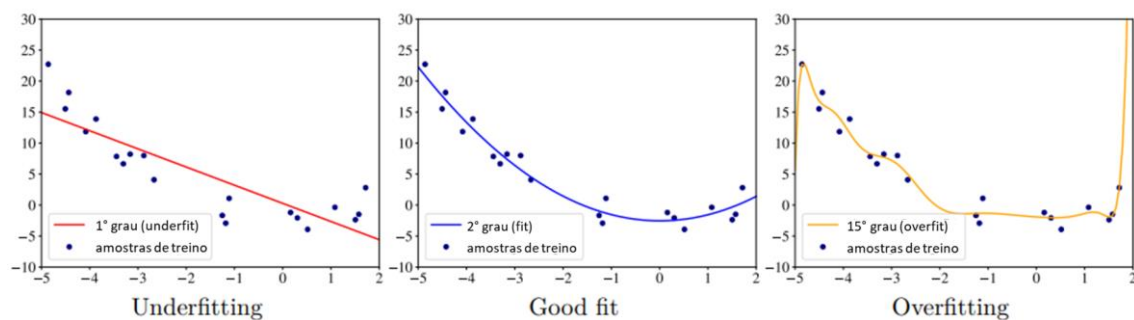


Figura 7: Exemplos de formas de curvas a depender do ajuste do dado. A - *Underfitting* para um modelo linear; B - Bom ajuste para um modelo quadrático; C - *Overfitting* para um modelo polinomial de grau 15 (Adaptado de Burkov, 2019).

3.2.2 Area Under the ROC Curve (AUC)

A curva ROC (“Receiver Operating Characteristic”) é um método comumente utilizado para avaliar o desempenho da classificação de modelos. Esse método de avaliação parte de uma combinação entre a taxa de verdadeiros positivos (TPR) e a taxa de falsos positivos (FPR) (Equação 1) para construção de uma imagem resumida da classificação do desempenho (Figura 8). O cálculo da TPR e da FPR é realizado a partir dos valores de Verdadeiros Positivos (TP), Falsos Positivos (FP), Verdadeiros Negativos (TN) e Falsos Negativos (FN).

$$TPR = \frac{TP}{(TP+FN)} \text{ e } FPR = \frac{FP}{(FP+TN)} \quad (1)$$

A construção da curva ROC é feita a partir dos valores de FPR (eixo X) e TPR (eixo Y), em que a conexão do valor limiar dos pontos gera a curva ROC. Quanto maior a área abaixo da curva ROC (AUC-ROC) melhor é o classificador. Um classificador que obtenha valor maior que 0,5 tem um desempenho melhor que um classificador aleatório, caso o resultado seja inferior a 0,5 existe algum problema em seu modelo. Um classificador perfeito terá um valor de 1.

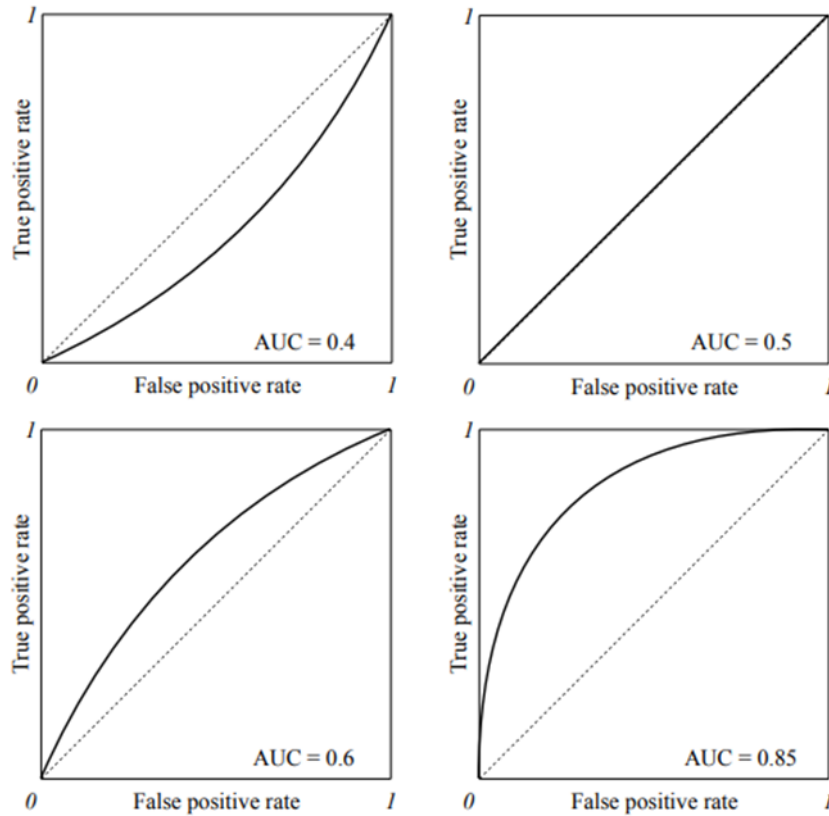


Figura 8: Exemplos de curvas para diferentes valores de AUC abaixo da ROC. Extraído de Burkov, 2019.

3.2.3 Acurácia (*Accuracy*)

A acurácia é dada pelo número de exemplos classificados corretamente dividido pelo número total de exemplos classificados. A acurácia é uma métrica útil quando os erros na previsão de todas as classes são importantes. Em termos da matriz de confusão, ela é dada pela equação 2:

$$Acurácia = \frac{TP+TN}{TP+TN+FP+FN} \quad (2)$$

3.2.4 Precisão (*Precision*)

Define a acurácia do modelo por meio da avaliação de quanto dos pontos amostrais correspondem as litologias previamente mapeadas. A métrica serve para avaliar o custo de falsos positivos (Equação 3).

$$Precisão = \frac{Verdadeiros\ Positivos}{Verdadeiros\ Positivos + Falsos\ Positivos} \quad (3)$$

3.2.5 Recall

Calcula quantos dos dados tidos como positivos são positivos verdadeiros, taxa de verdadeiros positivos. Essa métrica é adequada para avaliar o custo associado a falsos positivos (Equação 4).

$$Recall = \frac{Verdadeiros\ Positivos}{Verdadeiros\ Positivos + Falsos\ Negativos} \quad (4)$$

3.2.6 F1-Score

Essa métrica combina os valores de Precisão e Recall (Equação 5) para mensurar se há um balanço entre positivos verdadeiros e falsos negativos.

$$F1_Score = 2 * \frac{Precisão * Recall}{Precisão + Recall} \quad (5)$$

3.2.7 Validação Cruzada

Em casos onde não se dispõe de um conjunto de validação adequado para o ajuste dos hiperparâmetros de um modelo, a validação cruzada (*cross-validation*) é uma técnica que pode ser utilizada. Em situações em que se possui poucos exemplos de treinamento, a separação do *dataset* em conjunto de validação e teste separados pode ser inviável. Nesse contexto, os dados serão divididos apenas em conjunto de treino e teste, e a validação cruzada é aplicada ao conjunto de treino para simular um conjunto validação.

A validação cruzada no seguinte formato: primeiramente os valores dos hiperparâmetros que se planeja avaliar são fixados. Feito esse ajuste, o conjunto de dados de treino é dividido em ‘k’ subconjuntos de tamanhos iguais, denominados *folds*.

Comumente, é utilizada a validação cruzada com *k-folds* onde 'k' é igual a 5. Dessa forma os dados de treino são divididos aleatoriamente em 5 *folds*, onde cada *fold* contém 20% dos dados de treino.

Completada essa etapa, 5 modelos vão ser testados, tendo como conjunto de validação os dados de uma *fold* e como dados de treinamentos todas as outras *folds* com exceção da que está se usando como validação. Esse processo é repetido cinco vezes, até que todas as *folds* tenham sido treinadas como um conjunto de validação, em cada modelo é avaliado desempenho e ao fim uma média com os cinco valores das métricas calculadas é fornecido.

4. Materiais e Métodos

Neste capítulo serão abordadas as características dos conjuntos de dados utilizados para realização da pesquisa. Uma ênfase maior é dada aos dados aerogeofísicos e sua rotina de processamento, uma vez que a informação extraída durante o processamento compõe uma parte significativa da base de dados utilizada. Para implementação do *Random Forest*, configurou-se o número de amostras e alguns parâmetros foram ajustados para a realização da classificação.

4.1. Aerogeofísica

Os produtos aerogeofísicos, aeromagnetométricos e aeroradiométricos, utilizados no projeto foram gerados a partir da base de dados pertencente ao aerolevante do Projeto Aerogeofísico Tocantins (CPRM, 2006), situado na porção sul do Estado do Tocantins, estendendo-se sobre uma pequena porção do norte do Estado do Goiás (Figura 9). O levantamento foi realizado a partir de um acordo de cooperação MMA (Ministério do Meio Ambiente) – ANP (Agência Nacional do Petróleo) – SBG-CPRM (Serviço Geológico do Brasil) que permitiu a execução do aerolevante do Projeto Tocantins através da empresa AeroGeoPhysica Latino América (AGP-LA).

4.1.1 Características do Projeto Aerogeofísico Tocantins

Duas equipes participaram do levantamento, utilizando aeronaves Cessna C208B *Grand Caravan* PT-MEB e PT-MEX, equipadas com magnetômetros de bombeamento ótico de vapor de césio da SCINTREX, modelo CS-3, com intervalos de leituras de 0,1 segundos e gamaespectrômetros da Pico Envirotec, modelo GRS 410, de 512 canais espectrais e sistema de navegação GPS, com receptor Javad Euro GD de dupla frequência e 12 canais, com medições radiométricas de 1,0 segundos.

Cada aeronave foi responsável por uma área de prospecção, subárea 1 e Subárea 2, respectivamente. A área no qual o projeto está inserido corresponde a Subárea 2 (Figura 9), a qual teve seu levantamento realizado pela aeronave PT-MEX entre 06/08/2006 e 01/01/2007, sendo 67.798,48 km de linhas de voo e 3.587,25 km de linhas de controle, contando com uma base de operações na cidade de Barreiras, no estado da Bahia. A declinação da área de estudo tem valor de aproximadamente $-20,76^{\circ}$ com uma inclinação de $18,09^{\circ}$, para o período do levantamento realizado pela aeronave PT-MEX (CPRM, 2006).

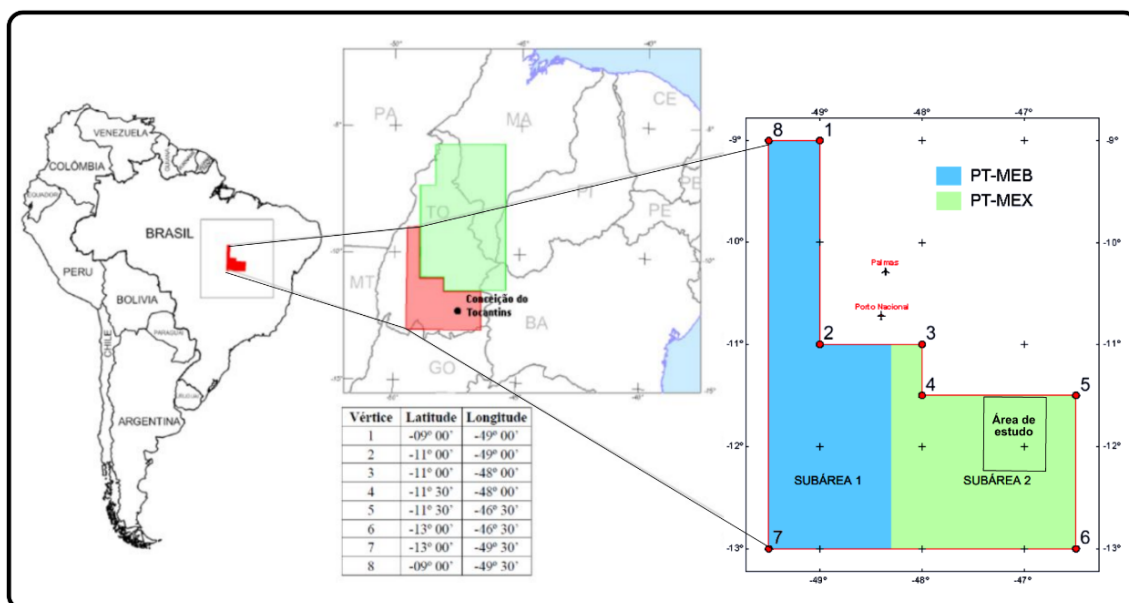


Figura 9: Localização da área do levantamento do Projeto Aerogeofísico Tocantins, em vermelho e a área de magnetometria e gamaespectrometria dos Levantamentos Aerogeofísicos da Bacia do Parnaíba em verde. Tabela contendo as coordenadas dos vértices do levantamento. Modificado de CPRM, 2006.

A etapa de aquisição de dados pelo Projeto Tocantins foi executada no período de 26 de julho de 2005 a 02 de janeiro de 2006. O levantamento aerogeofísico recobriu 164.003,34km de perfis aeromagnetométricos e aerogamaespectrométricos de alta resolução, com linhas de voo e de controle espaçadas de 0,5 km e 10 km, orientadas nas direções N-S e E-W, respectivamente. A velocidade média de voo foi de 270 km/h, a altura de voo foi fixada em 100m sobre o terreno (CPRM, 2006).

4.1.2 Gamaespectrometria

A gamaespectrometria é um método geofísico que visa a detecção e quantificação de radioatividade, a partir da emissão natural de raios gama (γ) procedentes das rochas (Dentith & Mudge, 2014).

A radiação gama pode ser emitida a partir da desintegração nuclear ou o decaimento radioativo de um átomo instável. Átomos de um mesmo elemento contendo diferentes números de nêutrons são denominados isótopos, sendo identificados pelo seu número de massa. Alguns isótopos são instáveis e mudam para um núcleo mais estável através da emissão de radiação, estes são chamados de isótopos radioativos ou “radioelementos”. O decaimento gama segue após o decaimento α ou β e não altera o número atômico ou o número de massa. A emissão de radiação γ está associada a uma redução na energia do isótopo filho de um estado excitado para um estado fundamental,

ou para um estado excitado de menor energia e nesse caso ocorrem mais emissões γ .

No planeta ocorrem mais de 50 elementos que naturalmente se comportam de maneira radioativa, no entanto, a radiação terrestre é dominada pelo produto da emissão de radioatividade de três elementos: potássio (K), tório (Th) e urânio (U). Os outros radioelementos são muito raros e/ou emitem uma radiação muito fraca, a ponto de ser de significância (Dentith & Mudge, 2014), de maneira que somente o K, Th e o U emitem raios gama suficientes para serem detectados e medidos em levantamentos aéreos (Minty, 1997).

A contagem total (CT) da radiação é mensurada a partir de um espectrômetro, com base em todos os raios gama na janela energética de 0,41 a 2,81 MeV. Nesse intervalo são caracterizados canais dos radioelementos onde os raios gama emitidos pelo potássio (^{40}K) tem um pico de energia a 1,46 MeV. Já os isótopos ^{238}U e ^{232}Th tem seus raios gama calculados através de seus produtos decorrentes do decaimento radioativo (^{214}Bi e ^{208}Tl , respectivamente) e tem seus raios gamas centrados em picos de energia de 1,76 e 2,61 MeV, respectivamente (Ribeiro, 2014). Dessa forma, as estimativas de concentração de U e Th são referidas como Urânio equivalente (eU) e Tório equivalente (eTh). As concentrações na crosta do K variam de 2 a 2,5%, a variação do U é de 2 a 3 ppm e a do Th é de 8 a 12 ppm (IAEA, 2003).

Analisando a variação de teores de silício entre rochas ácidas a ultrabásicas, nota-se que o aumento na concentração de radioelementos é correlacionado com o aumento de Sílica (Si) na rocha (Dickson & Scott, 1997). A assinatura dos litotipos pode variar a depender do tipo de composição da rocha e estilo de intemperismo que influenciará os processos de alteração que a rocha irá ser submetida. Alguns fatores a ser considerado para a análise de grupos litológicos pela assinatura radiométrica, são os solos transportados que mascaram por muitas vezes a assinatura radiométrica das rochas e a sobreposição de concentração de radioelementos para diferentes tipos de rocha e solo (Figura 10).

Alguns fatores podem influenciar os valores das medidas de radiação, entre eles: cobertura de solo (alcance médio de 30 a 40 cm); vegetação; umidade do solo; variações de temperatura atmosférica; umidade relativa e presença de nuvens; fenômeno de inversão térmica, que pode impedir a dissipação do radônio livre na atmosfera gerando falseamento dos dados; grande variação topográfica; efeitos direcionais; e as características do aerolevanteamento (Ribeiro, 2014).

Tipo de Rocha	Rocha			Solo		
	K %	U ppm	Th ppm	K %	U ppm	Th ppm
Intrusivas						
Granitoides	0.3–4.5 (2.4)	0.4–7.8 (3.3)	2.3–45 (16)	0.4–3.9 (2.1)	0.5–7.8 (2.7)	2–37 (13)
Rocha Gnaissica	2.4–3.8 (2.4)	2.1–3.6 (2.5)	18–55 (15)	0.7–1.9 (1.3)	1.6–3.8 (2.2)	6–19 (12)
Pegmatito	2.6–5.5 (3.7)	0.3–1 (0.7)	0.3–9.6 (2)			
Aplito	0.6–4 (2.4)	1–8 (3.3)	3–20 (7)			
Quartzo-Feldspato Porfiritico	1–5 (2.9)	1.3–2.9 (1.7)	6–14 (13)			
Intrusivas Intermediárias	0.7–5.6 (2.7)	0.1–1.2 (0.8)	0.8–6.1 (2.4)	0.7–3.4 (1.6)	1.5–2.3 (1.9)	2.9–8.4 (5.6)
Intrusivas Máficas	0.1–0.8 (0.4)	0.0–1.1 (0.3)	0.0–3.1 (1.2)			
Extrusivas						
Vulcânicas Félsicas	2.0–4.4 (3.7)	1.4–13 (2.4)	13–28 (17)	1.8–3.2 (2.4)	1.3–2.4 (2.1)	10–18 (13)
Vulcânicas Intermediárias	1.8–4.1 (2.7)	0.9–5.6 (2.3)	1.5–15. (9)	1.0–2.7 (1.9)	1.2–3.6 (2.1)	4–17 (10)
Andesitos de Baixo-K	0.7–0.9 (0.8)	1.0–2.5 (1.6)	3–8 (5)	0.8–1.5 (1.1)	1.2–1.5 (1.3)	4–6 (5)
Vulcânicas Máficas	0.3–1.3 (0.9)	0.3–1.3 (0.7)	2.0–5.0 (3.0)	0.2–1.4 (0.7)	0.6–2.5 (1.6)	3.3–13 (7.9)
Vulcânicas Ultramáficas	0.2–0.9 (0.4)	0.3–0.9 (0.6)	0.0–4.0 (1.2)	0.6	2.0	6.
Rochas Sedimentares						
Folhelhos Arqueanos	0.4–1.6 (0.9)	0.3–1.3 (0.9)	1–5 (2.7)	0.8	1.2	3
Outros Folhelhos	0.1–4.0 (2.6)	1.6–3.8 (2.6)	10–55 (19)	0.7–3.0 (1.5)	1.2–5 (2.3)	6–19 (13)
Arenitos	0.0–5.5 (1.8)	0.7–5.1 (2.3)	4–22 (12)	0.1–2.4 (1.3)	1.2–4.4 (2.1)	7–18 (11)
Carbonatos	0.0–0.5 (0.2)	0.4–2.9 (1.6)	0–2.9 (1.4)			

Figura 10: Tabela com teores de concentração de K, U e Th das principais rochas e solos da Australia. (Adaptado de Dickson e Scott, 1997).

4.1.2.1. Processamento de Dados Gamaespectométricos – Técnicas de realce

O pré-processamento dos dados gamaespectométricos começou na análise dos dados, via parâmetros estatísticos como a média, mediana, e valores máximos e mínimos para os canais de K (%), eTh (ppm), eU (ppm) e Contagem Total (ppm) para área na totalidade e isoladamente para as principais unidades geológicas.

A partir desta análise inicial notou-se a presença de teores negativos no banco de dados, representando uma inconsistência física para uma fonte radioativa. Para a correção desses valores foi aplicada uma normalização (Equação 6) com o intuito de não mais haver valores negativos e zeros, uma vez que razões seriam realizadas posteriormente (Mendes *et al.*, no prelo, 2023).

$$X_{normalized} = \frac{(X - X_{min})}{(X_{max} - X_{min})} + 0.001 \quad (6)$$

Realizada a correção no Banco de Dados foram elaborados *Grid's*, com o uso do *software OasisMontaj*, inicialmente para os três canais de radioelementos e o canal de contagem total (Figura 11), utilizando o interpolador Mínima Curvatura (Briggs (1974) e

Swain (1976)) para células de 125 metros, $\frac{1}{4}$ do espaçamento das linhas de voo. A partir dos Grid's gerados para os três canais foi gerada uma composição ternária utilizando o padrão de cor RGB, composição de potássio, tório e urânio (Figura 12 – A).

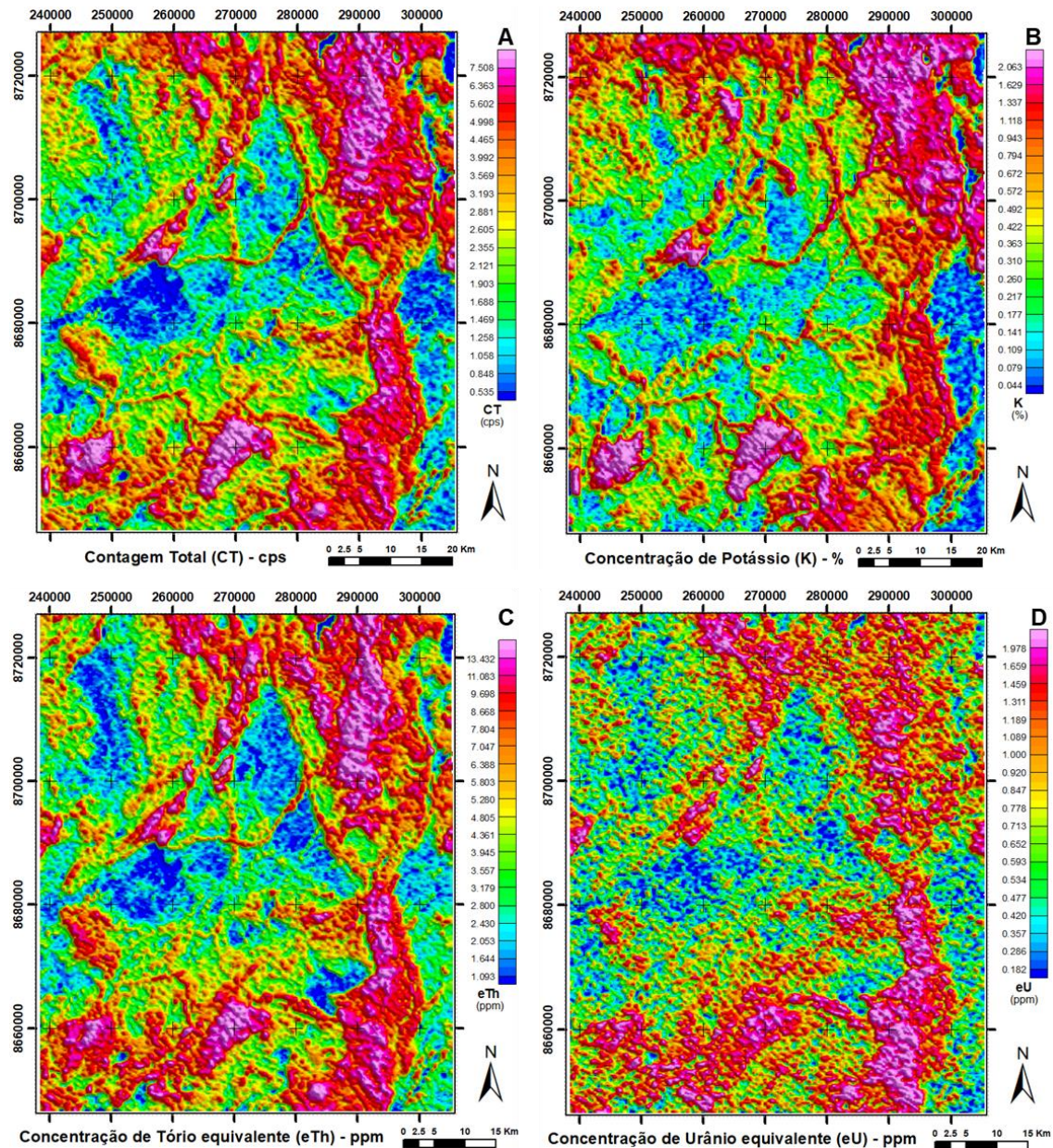


Figura 11: Imagens geradas a partir do processamento dos dados radiométricos, imagens sombreadas: A – Contagem Total (CT); B – Concentração de Potássio (K); C – Concentração de Tório Equivalente (eTh); D – Concentração de Urânio Equivalente (eU).

A composição ternária é uma das ferramentas mais utilizadas para realização de análise qualitativa do terreno, mediante que as variações e combinações de cores implicam valores relativos para concentração de radioelementos para a célula interpolada. Este é um produto que por muitas vezes, a depender da resolução do levantamento, traz informações de grande valia para confecção de mapas geológicos, uma vez que contatos

entre alguns tipos de unidade apresentam um forte contraste.

Para análise quantitativa dos radioelementos, foram gerados alguns produtos estatísticos com o intuito de melhor caracterizar a assinatura radiométrica da área de estudo em geral, e para análise pontual foram recortados os pontos amostrados pelo aerolevantamento somente para cada uma das unidades que ocorrem na área de estudo. Os dados foram analisados e comparados com valores de referência na literatura, depois dispostos em diagramas do tipo *Box & Whiskers* para uma melhor visualização.

Com o intuito de gerar produtos que realçassem determinadas unidades na área, visando a utilização de seus valores na implementação da classificação supervisionada, foi realizada uma série de raízes e normalizações entre os canais de radioelementos.

4.1.2.2 Parâmetro F

Outro produto que evidencia o enriquecimento de potássio e urânio em relação ao tório é o Parâmetro F (Fator F) (Figura 12 – B), descrito por Efimov (1978 *apud* Gnojek e Prichystal, 1985). Esta razão (Equação 7) é utilizada em ambientes geológicos em que há ocorrência de processos hidrotermais, como, por exemplo, alterações fílicas e potássicas que tem como um resultante um enriquecimento de potássio no sistema (Dickson e Scott, 1997).

$$F = \frac{K*U}{Th} = \frac{U}{\frac{Th}{K}} = \frac{K}{\frac{Th}{U}} \quad (7)$$

Esta razão coloca em evidência áreas abundantes em potássio e urânio devido à mobilidade destes radioelementos, o urânio por ter uma alta mobilidade tende a estar também presente em processos hidrotermais, já o tório que tem como característica a baixa mobilidade, em ambientes hidrotermais apresenta uma baixa concentração. É importante ressaltar que áreas com forte intemperismo e alta lixiviação tendem a apresentar uma baixa concentração, enquanto maciços submetidos a processos hidrotermais apresentam alta concentração.

Poucos estudos trazem valores de referência para tipos de rocha e suas respectivas concentrações do Parâmetro F, Efimov (1978 *apud* Gnojek e Prichystal, 1985) indica que valores em torno de 1,2% a 1,3% são comuns em rochas não alteradas em quanto o valor de referência para rochas alteradas gira em torno de 2 – 5%, podendo haver casos em que este valor é superior, no entanto, esses valores são relacionados a rochas no hemisfério

norte e podem não ter correspondência as rochas da área de estudo.

4.1.2.3 Potássio Anômalo (Kd) e Urânio Anômalo (Ud)

Quando medido os teores dos radioelementos temos fatores primários que influenciam seus valores como a variação de litologia superficial e tipos de solo, bem como as condições ambientais. Fatores secundários como acúmulo de determinados minerais e segregação de fluidos por resultado de frentes de alteração hidrotermal. Esta metodologia vem com este intuito de reduzir os efeitos causados por variações primárias para uma melhor compreensão dos processos secundários (Pires, 1995).

Para Saunders *et al.* (1987, 1993), os valores ideais de urânio e potássio podem ser definidos para cada amostra tendo como base os teores de tório, uma vez que os efeitos litológicos e ambientais que afetam a concentração de tório afetam também os teores de potássio e urânio de modo previsível. As concentrações de Tório podem ser utilizadas para uma melhor estimativa dos valores ideais de potássio e urânio, pelo fato de o mesmo ser um elemento com baixa mobilidade quando comparado ao urânio (Adams & Gasparini, 1970), e por apresentar uma relação de concentração frente ao potássio e ao urânio que corresponde a uma função linear que passa pela origem (Pires, 1995).

Para o cálculo dos valores ideais de K e U foi proposta por Saunders *et al.*, 1987, as seguintes equações para a normalização dos teores de potássio (Equação 8) e urânio (Equação 9) em função do Th.

$$K_i = \left(\frac{\text{média de } K_s}{\text{média de } Th_s} \right) * Th_s \quad (8)$$

$$U_i = \left(\frac{\text{média de } U_s}{\text{média de } Th_s} \right) * Th_s \quad (9)$$

Em que:

K_s, **Th_s** e **U_s** = valores aferidos;

K_i e **U_i** = valores ideais obtidos a partir da normalização em relação ao Th.

As equações 10 e 11 representam os valores ideais de potássio (K_i) e Urânio (U_i), para calcular os desvios dos valores reais com base nos valores ideais, utilizam-se as equações 10 e 11 abaixo:

$$K_d = \frac{(K_s - K_i)}{K_i} \quad (10)$$

$$U_d = \frac{(U_s - U_i)}{U_i} \quad (11)$$

Em que: K_d , e U_d = desvios dos valores ideais para potássio (Figura 12 – C) e urânio, respectivamente.

4.1.2.4 Índice Máfico (IM)

A técnica denominada Índice Máfico tem o objetivo de individualizar corpos máficos e ultramáficos (Pires & Moraes, 2006) uma vez que a influência do material rico em ferro presente na superfície, como, por exemplo, solos lateríticos, é eliminado. O produto relaciona os dados magnéticos e radiométricos por meio da razão entre a Amplitude do Sinal Analítico e o produto dos três canais de radioelementos (K, eTh e eU) através da Equação 12 foi gerado (Figura 12 – D).

$$IM = \frac{ASA}{K * eTh * eU} \quad (12)$$

4.1.3 Magnetometria

O método magnetométrico, ou magnético, baseia-se na variação do campo magnético terrestre, estas variações estão dispostas nos produtos aeromagnéticos como anomalias, no qual são interpretadas como resultado da fossilização de rochas magnéticas (e.g., Yandjimain *et al.*, 2018) na crosta terrestre. Diante disso, a quantidade, composição e modo de distribuição da mineralogia magnética na rocha são variáveis responsáveis pela variação do campo magnético terrestre (Luiz & Silva, 1995).

O campo magnético terrestre é definido pela magnitude (F), inclinação (I), que corresponde ao ângulo entre a superfície da Terra e o vetor F, e declinação (D), ângulo entre a componente horizontal do vetor F e o norte geográfico. Na superfície da Terra é medido um campo magnético resultante do campo magnético interno, oriundo da porção líquida do núcleo da Terra, do campo magnético externo, resultado das flutuações magnéticas causadas por correntes elétricas produzidas na ionosfera e tempestades solares, e do campo magnético anômalo proveniente da magnetização dos minerais constituintes da rocha (Dentith & Mudge, 2014).

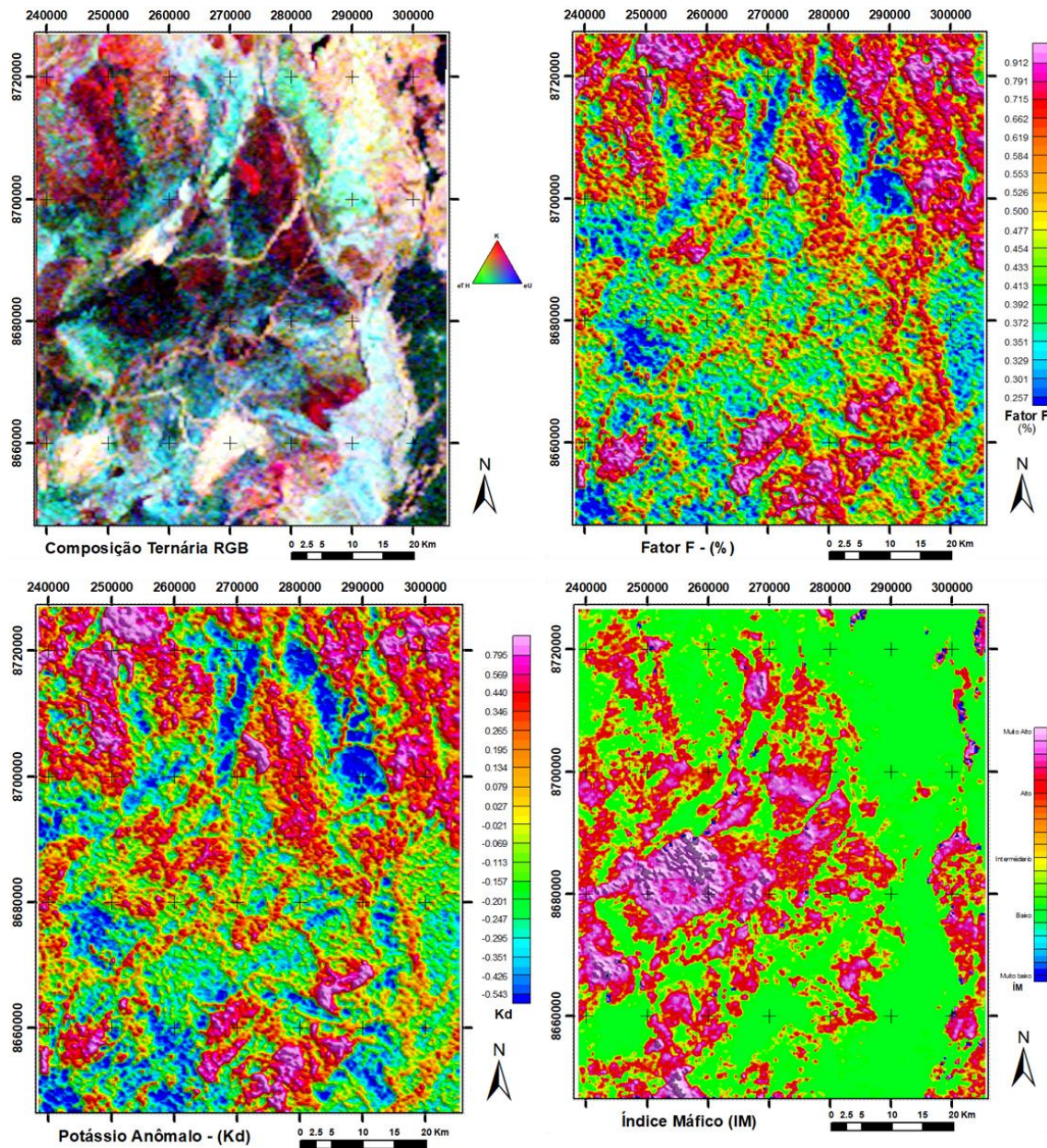


Figura 12: Imagens geradas a partir do processamento dos dados radiométricos, imagens sombreadas: A – Composição Ternária (Red: K, Green: eTh, Blue: eU); B – Fator F (%); C –Potássio Anômalo; D – Índice Máfico.

As rochas podem apresentar dois principais comportamentos sob a influência do campo magnético: magnetização induzida, proveniente da aplicação de um campo magnético ambiente, e a magnetização remanente ou remanescente, que existe independente da aplicação de um campo externo e ocorre em substâncias ferri- e ferromagnéticas, caracterizadas por um alinhamento natural do momento magnético. Os materiais podem ser classificados em três grupos com base em suas propriedades magnéticas, sendo classificados como diamagnéticos, paramagnéticos e ferromagnéticos (Dentith & Mudge, 2014).

A magnetometria aérea é muito utilizada para a diferenciação de domínios litológicos, dado que diferentes rochas vão possuir diferentes assinaturas magnéticas. A propriedade física relacionada à magnetometria é a susceptibilidade magnética, que possui um comportamento amplo nas rochas, como visto na Figura 13. Os valores de susceptibilidade se sobrepõem em diferentes tipos de rochas, a ponto de a diferenciação de cada unidade ser feita com base em uma análise conjunta de propriedades físicas (Dentith & Mudge, 2014).

4.1.3.1 Processamento de Dados Magnetométricos – Técnicas de Realce

O dado magnético obtido pelo Projeto Aerogeofísico Tocantins é chamado de Campo Magnético Total (CMT), que corresponde ao produto de todos os campos magnéticos atuantes naquele ponto, por exemplo, Campo Externo, Campo Interno e Campo Crustal. A etapa de pré-processamento (redução de dados) realizada pela empresa AGP-LA, utilizando o software *Geosoft Montaj* 5.1.8 (Reeves, 2005) no qual foram realizadas ações utilizando parâmetros de nivelamento e filtragem no dado referente ao CMT, sendo estas ações: correção do erro de paralaxe, compensação magnética, remoção da variação diurna, nivelamento dos perfis, micronivelamento dos perfis e a remoção do IGRF (*International Geomagnetic Reference Field*) (CPRM, 2006). Após a rotina de processamento, o campo residual gerado pelas correções e filtros é chamado de *Total Magnetic Intensity* (TMI) (Figura 14 - A) é obtido, e tem como unidade de medida nanoTesla (nT).

A interpolação dos dados foi realizada em malha regular de 125 metros x 125 metros por meio do interpolador bi-bidirecional (*Bi|Grid*), o primeiro produto interpolado foi o TMI (*Total Magnetic Intensity*). A partir do produto TMI foram executados procedimentos com o auxílio do software *Geosoft* v.8.1 e gerados uma série de realces e filtros como: Dx, Dy, Dz (Derivadas direcionais em X, Y e Z), ASA (Amplitude do Sinal Analítico), GHT (Gradiente Horizontal Total) e a Derivada TILT.

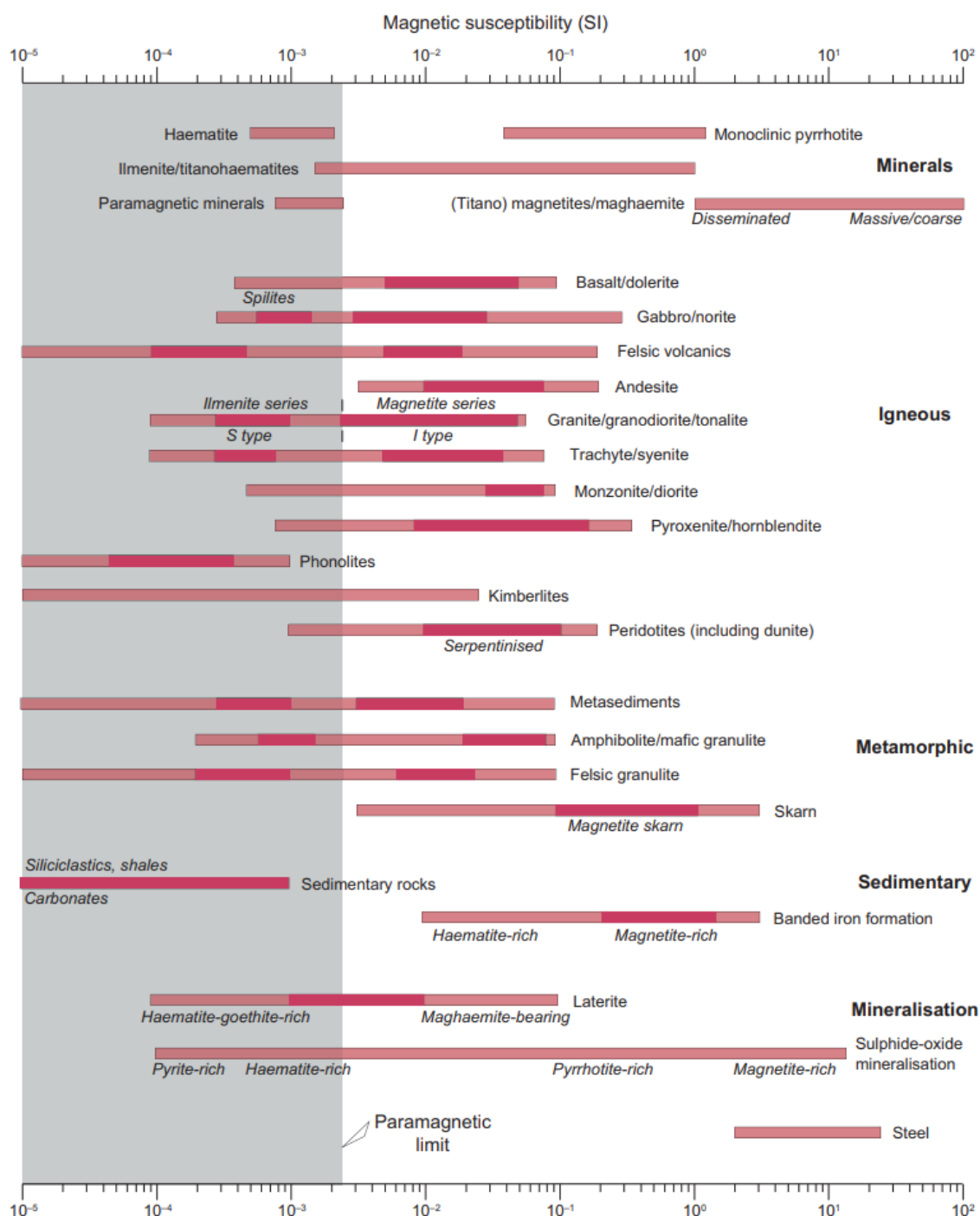


Figura 13: Intervalos de susceptibilidade magnética para diferentes minerais e tipos de rocha. As porções de coloração mais forte indicam os intervalos mais comuns. Retirado e modificado de Dentith & Mudge, 2014.

4.1.3.2 Derivadas Direcionais (Dx, Dy e Dz)

As derivadas tendem a “aguçar” bordas das anomalias e realçar feições rasas que dizem respeito aos menores comprimentos de onda e as mais altas frequências (Telford *et al.*, 1990), e é uma ferramenta também utilizada no processamento de dados

gravimétricos.

As derivadas em x e y, realçam feições somente nos planos horizontais, auxiliando na delimitação da extremidade de fontes pelo gradiente da propriedade física. As derivadas horizontais são mais comumente utilizadas para elaboração de produtos, onde a combinação destas traz uma melhor visualização dos contrastes laterais, tal como o Gradiente Horizontal Total (Grauch e Cordell, 1987).

A derivada vertical é a mais comumente utilizada, pelo fato de ser um filtro passa alta que fornece uma imagem mais clara das estruturas geológicas. A primeira derivada vertical não possui um viés direcional, o que é chave para melhorar a resolução espacial e estruturas presentes no produto. Dessa forma, a Dz (Figura 14 – D) corresponde ao gradiente vertical, na direção do eixo z, sendo utilizada para a determinação/delimitação de feições rasas como: lineamentos, descontinuidades e contatos.

4.1.3.3 Amplitude do Sinal Analítico (ASA)

A imagem da Amplitude do Sinal Analítico (ASA) é obtida a partir do conjunto das derivadas direcionais, extraídas do campo magnético anômalo, sendo gerado pela Equação 13.

$$ASA = \sqrt{\left(\frac{dH}{dx}\right)^2 + \left(\frac{dH}{dy}\right)^2 + \left(\frac{dH}{dz}\right)^2} \quad (13)$$

O dado de AGT auxilia na melhor compreensão da geometria da fonte magnética e é independente da direção da magnetização de sua fonte dependendo da intensidade do campo (Milligan & Gunn, 1997; Dentith & Mudge, 2014), evidenciando as bordas de corpos largos e centralizando corpos estreitos (Blum *et al.*, 1999). Por posicionar a fonte em sua máxima amplitude, é utilizado como uma boa ferramenta para delimitação de contatos geológicos (Figura 14 - B).

4.1.3.4 Gradiente Horizontal Total (GHT)

O método do gradiente horizontal total foi primeiramente descrito por Cordell em 1979, para a delimitação de contatos a partir de dados magnéticos e gravimétricos. O método é baseado no princípio que contatos marcados por falhas quase verticais, são produzem uma anomalia de gravidade (contraste) no qual o gradiente horizontal é maior

nas bordas do que sobre o centro da anomalia (Grauch e Cordell, 1987), e a imagem (Figura 14 – C) é gerada a partir da equação 14 (Blakely, 1996).

$$GHT = \sqrt{\left(\frac{dH}{dx}\right)^2 + \left(\frac{dH}{dy}\right)^2} \quad (14)$$

O produto do método posiciona os picos das anomalias nas bordas e os valores mínimos no centro, de maneira que a amplitude do sinal diminui com o aumento da profundidade (Ferreira *et al.*, 2013).

4.1.3.5 Derivada *Tilt*

O método foi desenvolvido por Miller e Singh em 1994 e corresponde ao resultado do arco tangente da razão entre a derivada vertical e o gradiente horizontal total, conforme a equação 15. Este produto também é conhecido como a Fase do Sinal Analítico, descrito por Blum em 1999, e permite estimar mergulho da fonte magnética e o contraste de susceptibilidade magnética, evidenciando lineamentos e foliações magnéticas além de equalizar a amplitude de fontes rasas e profundas (Miller & Singh, 1994).

$$TILT = \tan^{-1} \left(\frac{Dz}{THDR} \right) \quad (15)$$

O resultado do produto é o ângulo tilt, não uma razão, que varia em um intervalo entre $-90^\circ < TILT < 90^\circ$, o ângulo terá seu valor máximo positivo no centro da fonte, o valor zero ou próximo de zero nas bordas da fonte, e valores negativos, fora da fonte magnética (Miller e Singh, 1994).

Um atributo da derivada *TILT*, que o diferencia dos demais filtros de detecção de bordas, é que o mesmo responde igual para fontes rasas e profundas, permitindo a visualização de fontes profundas que em outros produtos são mascarados pela presença de anomalias de alta frequência (Miller e Singh, 1994; Verduzco, 2004).

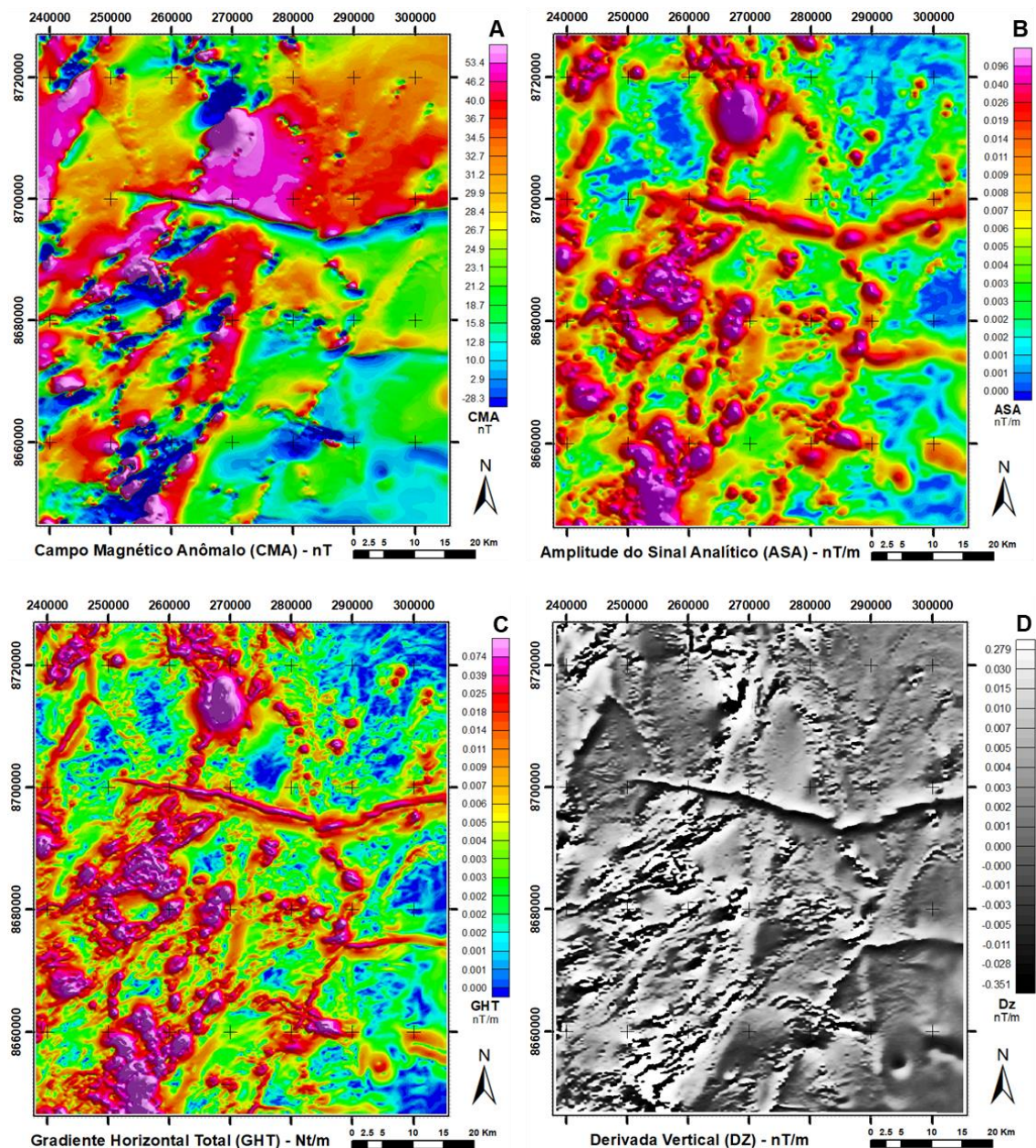


Figura 14: Imagens geradas a partir do processamento dos dados magnetométricos, imagens sombreadas: A – Total Magnetic Intensity (TMI); B – Amplitude do Sinal Analítico (ASA); C – Gradiente Horizontal Total (GHT); D – Derivada Vertical (DZ).

4.1.3.6 Inversão dos Vetores de Magnetização (MVI)

A utilização de métodos de inversão tem se tornado cada vez mais comum para interpretação de dados magnéticos, a modelagem de propriedades magnéticas das rochas a partir de observações do campo magnético vem sendo utilizada a anos no setor de exploração mineral. No entanto, essas abordagens por muito tempo eram limitadas por assumir que a magnetização da rocha era predominantemente induzida, de modo que a direção de magnetização estaria alinhada com o campo geomagnético (Macleod & Ellis,

2015).

Atualmente é de conhecimento que áreas de baixas latitudes magnéticas são tipicamente complexas devido à natureza do campo magnético. O método mais usual para simplificar a compreensão das anomalias magnéticas é a Redução ao Polo (RTP) (Baranov, 1957; Bhattacharya, 1965), porém o método se torna instável em baixas latitudes magnéticas e não apresenta bons resultados em locais onde há uma magnetização remanente significativa. A magnetização remanente tem um efeito bem maior do que se pensava, de forma que é um fator essencial para interpretação do campo magnético (McEnroe *et al.*, 2009).

Por este motivo, modelos de inversão de dados magnéticos convencionais, inversão por susceptibilidade, tem se mostrado não muito precisos, apresentando um caráter difuso e por vezes não centralizando bem as anomalias (Johnson & Aisengart, 2014). Isso ocorre, pois o método considera somente a magnetização induzida e não a magnetização remanescente. Quando a magnetização remanescente não é considerada, as anomalias não são centralizadas e as dimensões de topo e base das anomalias são subestimadas e/ou deslocadas (Figura 15) (Aisengart, 2013).

Um dos indicativos de remanência em áreas de baixas latitudes magnéticas é a presença de dipolos invertidos no TMI (Figura 14 – A), que no hemisfério sul tem como característica o polo negativo para norte e o positivo para sul (Blakely, 1996). Pelo fato de a área de estudo apresentar regiões com presença de remanência, além de estar localizada na Anômala Magnética do Atlântico Sul caracterizada por baixas latitudes magnéticas (Chapman & Bartels, 1940), a inversão dos vetores de magnetização apresenta resultados mais coerentes.

O MVI é um método que visa incorporar a magnetização induzida e remanente sem que haja a necessidade de um conhecimento prévio da magnitude ou direção do campo remanente (Ellis *et al.*, 2012; Macleod & Ellis, 2013; Barbosa, 2013; Johnson & Aisengart, 2014, 2015). O método utilizado na pesquisa utilizou-se de um algoritmo, integrado ao *software Oasis Montaj*, a partir do ambiente *VOXI Earth Modelling System* responsável por gerar modelos 3D de distribuição de susceptibilidade e dos vetores de magnetização.

O algoritmo do MVI visa a medição do campo magnético '**B**' em uma série de localidade '**r**' gera um volume **V** resultando em uma magnetização '**M**', resultando na equação direta para o Vetor de Magnetização (Ellis, *et al.*, 2012), conforme a Equação 16:

$$\mathbf{B}(\mathbf{r}_j) = \nabla \int_V \mathbf{M}(\mathbf{r}) \cdot \nabla \frac{1}{|\mathbf{r} - \mathbf{r}_j|} d\mathbf{r}^3 \quad (16)$$

A equação acima mostra que o vetor de magnetização $\mathbf{M}(\mathbf{r})$ é o parâmetro a ser recuperado pela inversão do campo magnético e pode ser reescrita de maneira simplificada, conforme a equação 17:

$$\mathbf{B} = G\mathbf{m} \quad (17)$$

O problema do MVI é resolver $\mathbf{m}$ (momento dipolo magnético), para um dado $\mathbf{B}$. Para que este problema inverso seja solucionado, é necessário que $\mathbf{B}$ seja submetido a condições de regularização. A implementação usada no algoritmo do *software Oasis Montaj* em seu menu VOXI, é através do regularizador de gradiente mínimo de Tikohonov, para resolver o problema magnético inverso do cálculo do vetor de magnetização minimizando a diferença do campo calculado e medido.

A inversão 3D MVI gera três modelos de componentes do campo magnético, amplitude (AMP), ePerpendicular (EPERP) eProjetada (EPROJ) (Aisengart, 2013; Ellis *et al.*, 2012). Um quarto produto gerado são os vetores de magnetização que correspondem aos vetores da componente Amplitude, em que cada ponto na área de estudo corresponde a amplitude magnética dividida por intensidade do campo induzido, resultando em um vetor sempre positivo (Martins & Vidotti, 2021).

A componente Projetada corresponde a projeção do vetor magnético na direção do campo induzido, podendo ter valores positivos, quando na mesma direção do campo induzido, e negativos quando em direção oposta. A componente Perpendicular é relacionada a amplitude perpendicular ao campo induzido e é sempre positiva, evidenciando anomalias do campo não induzido (remanência, demagnetização e anisotropia).

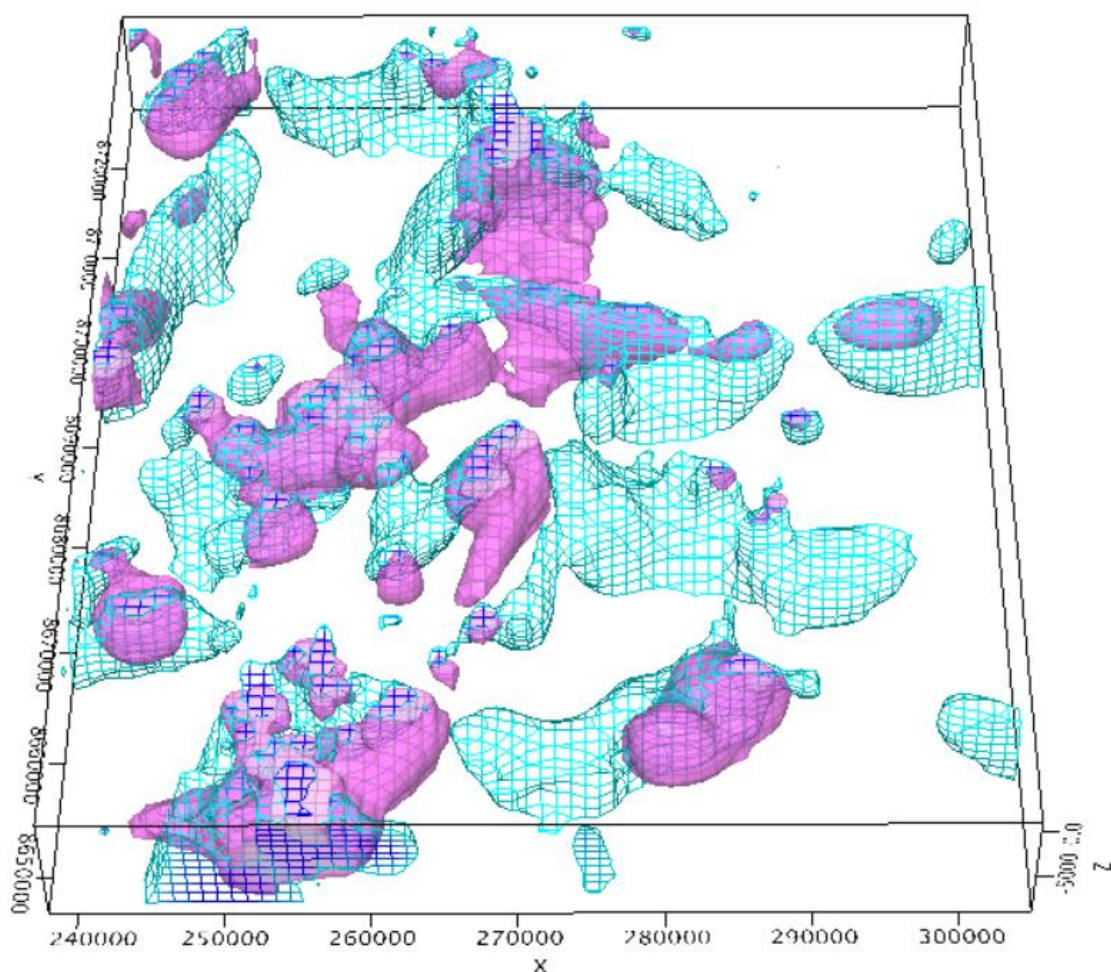


Figura 15: Comparação dos métodos de inversão de susceptibilidade e MVI, com valores de susceptibilidade equivalentes. Em azul, inversão convencional de susceptibilidade e em rosa, inversão dos vetores de magnetização (MVI). O MVI apresenta uma melhor centralização das anomalias, enquanto a inversão por susceptibilidade traz uma geometria mais alongada das anomalias.

4.2 Criação e elaboração da Base de dados para implementação do algoritmo *Random Forest*.

Inicialmente, foram gerados os produtos aerogeofísicos descritos anteriormente, com o uso do *software OasisMontaj*. Após a confecção dos GRID's (imagens aerogeofísicas), extraiu-se o valor do pixel a cada 250 metros, gerando uma malha regular com tamanho 66 km x 80 km (5.293 km²), visando reduzir o número de amostras e ajustar a base de dados para a mesma escala do mapa litológico de referência. Esse procedimento foi realizado para os dados magnetométricos e gamaespectrométricos, que haviam sido interpolados para células de 125 m pelos interpoladores *bi-directional* (Hinze et al., 2013, Isles and Rankin, 2013) e Mínima Curvatura (Briggs (1974) e Swain (1976)), respectivamente. Realizada essa operação exportou-se o *database* para o *Software ArcMap* para que valores de altimetria do satélite ALOS sensor Palsar fossem extraídos

para cada ponto da malha. Um procedimento semelhante foi realizado no mapa geológico da área de estudo (Figura 3), aplicou-se a ferramenta SpatialJoin do *software ArcMap* (Goodchild, 2011) para que em cada ponto da malha gerada fosse atribuída uma respectiva classe litológica (rótulo) correspondente à posição do ponto no mapa geológico.

Nesse trabalho, foi elaborado um mapa litológico “extrapolado” de maneira que os mapas geológicos existentes na área de estudo fossem analisados. A partir dessa análise, surgiu a necessidade de padronizar nomes de unidades geológicas e fechar contatos que não se concordavam na região limítrofe das cartas geológicas. Para menos classes serem alcançadas quando possível, agruparam-se classes com litotipos semelhantes, pertencentes à mesma unidade geológica.

A confecção do mapa litológico utilizado durante as etapas subsequentes teve como base as Folhas Dianópolis (Saboia *et al.*, 2019) e Arraias (Abdallah *et al.*, 2017), bem como o mapa geológico elaborado por Saboia (2021) por diferenciar em um maior detalhe as unidades pertencentes ao embasamento granítico-gnáissico, e por fim o mapa geológico do estado do Tocantins elaborado pelo Serviço Geológico Brasileiro (2022), uma vez que nele novos corpos pertencentes a Suíte Gameleira foram incorporados no mapa. O produto final desta etapa foi o mapa litológico generalizado da área de estudo, com o total de 17 classes litológicas extraídas do mapa geológico, com um total de pontos de 84.433, a distribuição de pontos amostrais por classe está indicada na Figura 16.

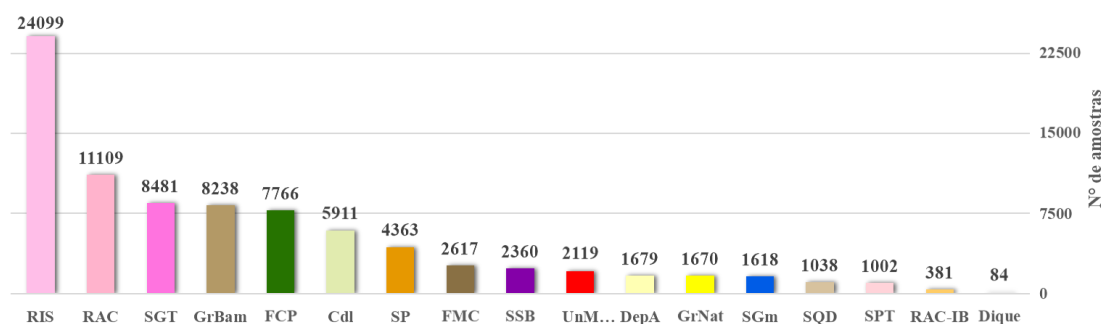


Figura 16: Diagrama indicando a proporção do número de amostras para cada uma das 17 classes litológicas. As cores das barras correspondem as mesmas cores do mapa geológico. O eixo X corresponde as seguintes unidades: Coberturas detrito-lateríticas (Cdl), Depósitos aluvionares (DepA), Diques (Dique), Gr. Bambui (GrBam), St. Gameleira (SGm), St. Granodiorítica aTonalítica (SGT), Un. Monzogranítica (UnM...), Gr. Natividade (GrNat), St. Peraluminosa (SP), St. Peraluminosa–tonalítica (SPT), Suíte Quartzodiorítica (SQD), St. Ribeirão Itaboca (RIS), St. Ribeirão das Areias (RAC), St. Ribeirão das Areias (RAC-IB), St. Serra Boqueirão (SSB), Fm. Morro do Carneiro (FMC) e Fm. Córrego do Paiol (FCP).

O banco de dados utilizado para a aplicação do *Machine Learning* conta com as seguintes colunas: Coordenadas (X, Y, Latitude e Longitude), Parâmetros radiométricos (CT, K, eTh, eU, Fator F, Ki, Kd, Ui e Ud), Parâmetros magnetométricos (CMA, ASA, GHT, slice de -100m do MVI), Outros (IM e Altimetria), e o *target* (Classe). Uma vez atualizado e com todos os campos presentes, este foi exportado e levado para o *Software Orange Data Mining* (Figura 17).

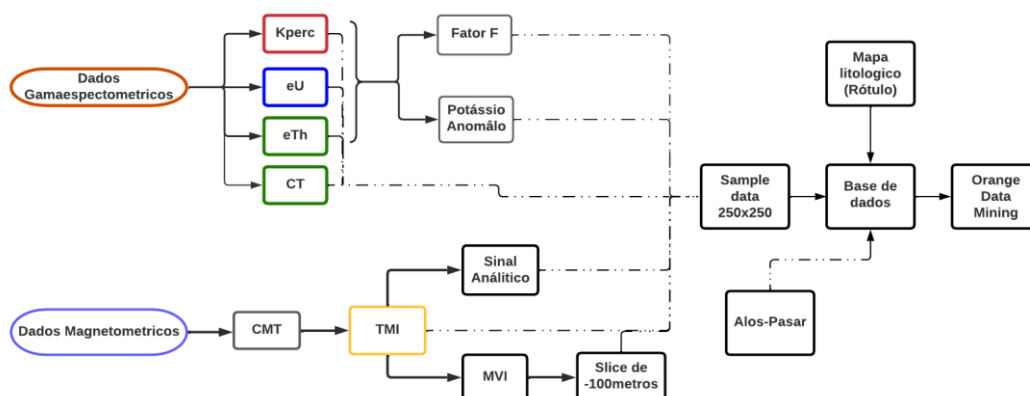


Figura 17: Fluxograma esquemático do processamento de dados aerogeofísico e composição da base de dados utilizada para implementação do Random Forest.

4.3 Análise de dados

Antes da implementação do algoritmo foi realizada uma análise exploratória dos dados, onde inicialmente avaliou-se a distribuição dos dados, como o *dataset* é muito extenso optou-se por uma análise utilizando diagrama QQ-Plot. Uma vez realizada a análise de normalidade, foram elaborados gráficos do tipo *Box&Whyskers* para os parâmetros K, eTh, eU e ASA. Essa etapa teve o intuito de analisar a distribuição de cada unidade para com os parâmetros e possíveis semelhanças entre classes, através da comparação de mediana, uma vez que os dados não são paramétricos e a média não é a melhor estimativa para o centro e dispersão dos dados.

A rotina de implementação de testes de modelos no *Orange* iniciou-se na preparação de amostras de treino e de teste, onde o tamanho da base de treino foi relativo a 100 amostras de cada classe, com exceção da classe “*Dykes*”, que possui somente 84 pontos (treinaram-se 70% do total de amostras). A base de treino compreendeu aproximadamente 2,0% dos dados totais, o que, conforme trabalhos como Cracknell *et al.*, (2014), Kuhn *et al.*, (2018) e Costa *et al.*, (2019) a percentagem utilizada foi 2,6%,

1,6% e 2,5%, respectivamente, o que demonstra que o número de amostras utilizado no presente trabalho é próximo aos encontrados na literatura.

A próxima etapa teve o objetivo de calcular qual seria o número de árvores a ser usado pelo algoritmo *Random Forest* para que a melhor acurácia e precisão fosse obtida para as predições, desta forma o modelo foi testado para os seguintes tamanhos de árvores: 100, 250, 500, 750, 1000 usando a técnica de validação cruzada com dez *folds*, predisposta no *widget Test & Score*. Foi realizada uma análise de importância dos parâmetros inseridos neste modelo teste, para que a partir do *index* Gini, fosse possível mapear as entradas de maior importância para o modelo.

Concluída essa etapa foram elaborados modelos variando parâmetros com base na importância que o algoritmo forneceu para cada parâmetro de entrada. Com o intuito de realizar predições com o intuito de encontrar a melhor combinação para rochas pertencentes a Suíte Gameleira.

Quando realizada a predição a tabela de dados foi exportada e manipulada em ambiente GIS, onde os pontos foram transformados para raster e posteriormente em polígonos para os dados fossem visualizados em formato de mapa.. Nessa etapa foram elaborados mapas preditivos com a mesma simbologia do mapa geológico, mapas probabilísticos indicando locais em que há uma classificação mais assertiva, e locais nos quais o algoritmo não consegue fornecer uma certeza.

Capítulo 5 – Resultados

Uso de *Random Forest* para a Detecção de Rochas Máficas e Ultramáficas da Suíte Gameleira, no Embasamento Arqueano Paleoproterozoico da Faixa Brasília Norte.

Luiz Fernando Cursino Tabosa¹, Elder Yokoyama¹, Thiago Lima Mendes¹, Pedro Maragno Almo¹, Guilherme Zakarewicz de Aguiar¹

¹ Universidade de Brasília, Brasil

Resumo

O embasamento do Orógeno Brasília em sua porção norte, corresponde a parte de um bloco crustal gerado no início do Paleoproterozoico, e é formado por um extenso volume magmatismos TTG (trondhjemitó-tonalito-granodiorito). A Suíte Gameleira, ocorre no embasamento da Faixa Brasília Norte (sudeste do estado do Tocantins), e corresponde a corpos predominantemente máficos com ocorrência subordinada de rochas ultramáficas no Domínio Almas-Conceição do Tocantins. A região é composta por variações de granitoides, sequências metavulcanossedimentares e metassedimentares. Dessa maneira, o objetivo dessa pesquisa é de usar um dos métodos de aprendizado de máquina para a detecção de ocorrências de possíveis novos corpos máfico-ultramáficos na Suíte Gameleira. Para isso, foram utilizados dados remotos, como levantamentos aerogeofísicos (magnetometria e radiometria), além de sensoriamento remoto. Foram gerados produtos derivados da aerogeofísica como Intensidade Magnética Total, Amplitude do Sinal Analítico, uma fatia em -100 metros de profundidade da Inversão dos Vetores de Magnetização, derivada vertical, derivada Tilt, os três canais radiométricos (K, eTh e eU), Contagem Total, Fator F e Potássio Anômalo. Com a informação desses produtos, foi gerada a base de dados e aplicado o algoritmo *Random Forest* (algoritmo de aprendizado de máquina supervisionado), onde foram treinados 1,96% da base de dados total (84535 amostras) para a geração do mapa preditivo. Para a verificação e avaliação dos dados preditos foram usadas três abordagens: uso de lineamentos magnéticos para averiguar um controle tectônico estrutural na colocação dos corpos, amostras de campo e o uso do MVI para a análise do comportamento magnético em profundidade dos corpos preditos, uma vez que os corpos da Suíte Gameleira apresentam remanência magnética. Concluiu-se na pesquisa que o uso desse método atrelado as formas de verificação utilizadas auxiliam na exploração de terrenos *greenfield* e *brownfield*, uma vez que o aprendizado de máquinas pode vir a aprimorar a cartografia existente, delimitando novas ocorrências ou mesmo sugerindo áreas em que se faz necessário melhor detalhamento. O uso de conhecimentos geológicos iniciais e outras informações não inseridas no modelo são de extrema importância, uma vez que o mapa preditivo não corresponde em si a verdade geológica e feita a predição é necessária que seja realizada interpretação do resultado por uma visão geológica.

Palavras-chave: Aerogeofísica: Faixa Brasília, Machine learning, Random Forest, Rochas Máficas Ultramáficas

Abstract

The northern part of the Brasília Orogen in its basement corresponds to a portion of a crustal block formed in the early Paleoproterozoic and consists of an extensive TTG magmatism (trondhjemite-tonalite-granodiorite). The Gameleira Suite is found in the basement of the North Brasília Belt (southeastern Tocantins) and consists of predominantly mafic rocks with subordinate occurrences of ultramafic rocks in the Almas-Conceição do Tocantins Domain. The region is composed of variations of granitoids, metavolcanosedimentary sequences, and metasedimentary rocks. The research aims to use a machine learning method for detecting potential new mafic-ultramafic occurrences in the Gameleira Suite. To achieve this, remote sensing data such as aerogeophysical surveys (magnetometry and radiometry) were used, along with remote sensing. Derived products from aerogeophysics were generated, including Total Magnetic Intensity, Analytic Signal Amplitude, a slice at -100 meters depth from the Magnetic Vector Inversion, vertical derivative, Tilt derivative, the three radiometric channels (K, eTh, and eU), Total Count, F-factor, and Potassium Anomaly. Using the information from these products, a database was created and the Random Forest algorithm (a supervised machine learning algorithm) was applied, training on 1.96% of the total database (84,535 samples) to generate the predictive map. Three approaches were used for data verification and evaluation: the use of magnetic lineaments to investigate structural tectonic control on the placement of bodies, field samples, and the use of MVI to analyze the magnetic behavior at depth of the predicted bodies, as the Gameleira Suite bodies exhibit magnetic remanence. The research concluded that the use of this method, combined with the verification methods used, assists in greenfield and brownfield exploration. Machine learning can enhance existing cartography by delineating new occurrences or suggesting areas that require further detailed mapping. The use of initial geological knowledge and other information not included in the model is crucial since the predictive map does not inherently represent geological truth, and interpretation of the results through a geological perspective is necessary.

Keywords: Aerogeophysics, Brazilian Belt, Machine learning, Random Forest, Ultramafic Mafic Rocks

1. Introdução

Com o decorrer do tempo e a evolução das ferramentas tecnológicas, as técnicas de mapeamento geológico vêm se atualizando de forma que o uso de técnicas indiretas para a aquisição de informações de maneira remota tem se tornado cada vez mais comum (Harris, 2015). O Brasil possui uma extensa área recoberta por aerolevantamentos de boa resolução, que são disponibilizados de maneira pública através do Serviço Geológico Brasileiro (SGB-CPRM), de forma que o uso desse tipo de dado no início de uma campanha prospectiva tem se tornado essencial.

Nessa linha de pensamento, com o advento da mineração de dados e a evolução dos métodos estatísticos o emprego de aprendizado de máquina (*Machine Learning*) vem se tornando uma ferramenta cada vez mais importantes no cenário prospectivo e na compreensão geológica de ambientes (Kuhn *et al.*, 2018, Kuhn *et al.*, 2019). Uma vez que a partir destas técnicas é possível a análise de dados geofísicos de alta resolução e em larga escala, que podem vir a gerar novas descobertas, um maior detalhamento cartográfico, complementando os métodos já existentes de mapeamento e prospecção (Schnitzler *et al.*, 2019).

A utilização dessas ferramentas tem ganhado cada vez mais importância dentro da mineração, devido ao seu caráter de baixo custo e ampla cobertura espacial que possibilita a avaliação de grandes áreas na busca de ocorrências minerais (Costa *et al.*, 2020, Hill *et al.*, 2021). Nesse cenário a busca por novas jazidas, e corpos que possam vir a ser prospectados de forma economicamente viável se torna cada vez mais necessária.

Sabendo disso, este trabalho tem foco no uso de ferramentas para a prospecção de novos corpos máfico-ultramáficos pertencentes a Suíte Gameleira (Correia Filho & Sá, 1980, CPRM, 2000, Saboia, 2021), no sudeste do Tocantins, entre os municípios de Conceição do Tocantins, Dianópolis e Almas. Até então, foram cartografados em torno de 12 corpos pertencentes a Suíte Gameleira (Saboia *et al.*, 2014, Abdalah *et al.*, 2015, Saboia *et al.*, 2019, Ribeiro *et al.*, 2022) com tamanhos variados, além de formatos semi-circulares a circulares. A maior intrusão pertencente a Suíte Gameleira foi estudada com um viés prospectivo pela CPRM, em 2000, onde foi realizado mapeamento geológico, análise de sedimentos de corrente e concentrados de bateia, que indicaram concentrações razoáveis de Cu-Co-Ni-Cr além de paládio, elemento pertencente aos PGE em canais de drenagem próximos a exposição da unidade.

A vista disso, os estudos prospectivos focados na Suíte Gameleira são poucos sendo eles: “Geologia e resultados prospectivos da Área Barra do Gameleira – Tocantins” (CPRM, 2000) e (Mendes, 2023), de forma que o potencial real desta unidade ainda não é claro. O presente trabalho, visa contribuir com a localização de novos corpos, que podem vir a ser foco de futuras pesquisas prospectivas ou mesmo no avanço da compreensão da gênese dessa unidade, bem como no avanço do conhecimento referente ao uso de aprendizado de máquina em geociências, por meio de testes elaborados a partir de diferentes dados de entrada e com o uso de ferramentas de verificação direta (dados de campo) e remotas que serão apresentadas no decorrer da pesquisa.

2. Geologia Regional

A área de estudo recobre cerca de 5.293 Km² e está localizada próxima as cidades de Almas, Dianópolis e Conceição do Tocantins, na porção centro sul do estado do Tocantins, Brasil. Situada no contexto da Província Tocantins (Almeida *et al.*, 1981, Fuck *et al.*, 2005, 2017), localizada na porção setentrional do cinturão orogenético Faixa Brasília, mais especificamente dentro do Domínio Almas Conceição do Tocantins (DACT), pertencente ao Bloco Crustal Cavalcante Natividade (BCCN) (Fuck *et al.*, 2014; Sousa *et al.*, 2016; Cordeiro *et al.*, 2014).

O DACT corresponde ao embasamento arqueano/paleoproterozoico, e é situado na porção nordeste do BCCN, é composto por um terreno granito-gnáissico com Suítes de idades Siderianas a Riagianas (Martins-Ferreira *et al.*, 2020; Saboia, 2020a, 2020b). Suítes estas que intrudem ou são colocadas lado a lado por contato tectônico com a sequência metavulcanossedimentar do Grupo Riachão do Ouro, dispostas em “faixas” de direção NS com ramificações para NE e NW envolvendo os domos graníticos-gnaissicos principalmente a norte na região próxima a Almas-Dianópolis (Fuck, 2014; Costa, 1985; Cruz & Kuyumjian, 1998; Borges *et al.*, 1998, 1999; Kuyumjian *et al.*, 2012).

O GRO corresponde a unidade mais antiga do DACT, com idade mínima de 2.47 -2.45 Ga com base na datação de granitoides da unidade Ribeirão das Areias que contem xenólitos de rochas do GRO (Cruz *et al.*, 2003; Martins-Ferreira *et al.*, 2017). Idades arqueanas paleoproterozoicas a partir de isótopos de Nd indicam uma idade TDM de 2.51 Ga para metabasaltos pertencentes a Formação Córrego do Paiol (Saboia & Meneguini, 2019). A Formação Córrego do Paiol é composta predominantemente de metabasaltos com raras ocorrências de rochas vulcânicas ultramáficas (Cruz & Kuyumjian, 1999;

Kuyumijan *et al.*, 2012). A Formação Morro do Carneiro é situada no topo da sequência metavulcanossedimentar do Grupo Riachão do Ouro e é formada por metassedimentos terrígenos e químicos. Cristais de zircão contidos dentro de uma rocha vulcânica félsica pertencente a Formação Morro do Carneiro, nas proximidades da cidade de Almas, datados a partir de U-Pb forneceram uma idade de 2206 ± 13 Ma (Dardenne *et al.*, 2009).

No presente trabalho a subdivisão de unidades para este terreno utilizada foi a proposta por Saboia (2021), em que o autor subdivide o terreno em unidades com base em estudos geocronológicos. As unidades de idade Sideriana correspondem ao Complexo Ribeirão das Areias (2.47 Ga) e a Suíte Ribeirão Itaboca (2.3 Ga) proposta por Saboia, 2021. Já as unidades de idades Riagianas que ocorrem dentro do DACT são: Unidade Monzogranítica (2.29 Ga), Suíte Quartzo – Diorítica (2.28 Ga), Suíte Granodiorítica a Tonalítica (2.26 Ga), Suíte Serra do Boqueirão (2.22 Ga) e Suíte Peraluminosa (2.2 Ga) (Cruz *et al.*, 2003; e referências nele contidas; Abdallah & Meneghini, 2017; Sousa *et al.*, 2016; Saboia, 2021).

As rochas da Suíte Gameleira estão em contato com as rochas do Complexo Almas-Cavalcante e do Grupo Riachão do Ouro de forma intrusiva ou por contato tectônico. Os corpos desta suíte correspondem a intrusões acamadadas, com formato semicircular a corpos alongados. Estes corpos foram submetidos, localmente, a um metamorfismo em fácies anfibolito a granulito (Correia Filho e Sá, 1980, CPRM, 2000) preservando em certos domínios estruturas ígneas primárias, é frequentemente observado substituições minerais características de um retrometamorfismo em fácies xisto-verde. Possuem assinaturas características de alta intensidade magnética e baixa resposta radiométrica.

A área de estudo encontra-se predominantemente dentro do DACT, onde a sudeste faz contato com rochas do Grupo Bambuí (Dardenne, 1978; Alvarenga, 2012), pertencente ao domínio Cratônico, e a norte ocorrem rochas do Grupo Natividade que foram depositadas acima do embasamento (Oliveira *et al.*, 2012; Praxedes, 2015; Saboia & Meneghini, 2019), as demais unidades se inserem no contexto do DACT (Fig. 1). Dentro do perímetro da área de estudo definida para este experimento, ocorrem 17 classes litológicas, em que algumas dizem respeito a subclasses dentro de uma unidade geológica, por se diferenciarem quando a composição mineralógica e litotipos, essas informações foram sintetizadas na Tabela 1.

Tabela 1: Síntese das 17 classes litológicas da área de estudo. Informações retiradas de Cruz & Kuyumjian, 1998, Kuyumjian *et al.*, 2012, Abdallah & Meneghini, 2017, Saboia & Meneghini, 2019, Saboia, 2021.

Unidade	Rótulo	Descrição
Coberturas detrítico-lateríticas	Cdl	Areia, argila, silte, horizontes de crosta laterítica
Depósitos aluvionares	DepA	Areia com intercalações de argila, cascalho e restos de matéria orgânica.
Diques	Dique	Diques máficos-ultramáficos, composto por Gabro, Peridotito, Piroxenito, Serpentinito. Foi descrito somente um afloramento, os demais foram inferidos.
Gr. Bambui	GrBam	(i) Formação Sete Lagoas – composta por argilitos laminados e siltitos finos na base, sotopostos por metacalcarenitos, metadolarenitos e metadolomitos micríticos e subordinadamente. Formação Serra de Santa Helena: Constituída por ardósias estratificadas e de cor acinzentada e metapelitos de composição silício-argilosa, exibindo clivagem ardósiana.
St. Gameleira	SGm	As rochas máficas, predominantes, são principalmente gabro, leucogabro, troctolito, olivina-gabronorito, olivina-gabro coronítico e norito. As rochas ultramáficas são representadas por dunitos, piroxenitos feldspáticos, peridotitos serpentinizados e serpentinitos.
St. Granodiorítica a Tonalítica	SGT	Litotipos predominantes dessa unidade são biotitas metagranodioritos e metatonalitos, com ocorrência secundária e hornblenda-biotita metatonalitos, biotita metamonzogranitos e quartzo metadiorito.
Unid. Monzogranítica	UnMonz	Composta principalmente por metagranitos de granulação média a grossa, inequigranular a porfírica. O corpo granítico Pau Ramalhudo (Correia Filho e Sá, 1980) é a intrusão mais característica desta unidade, orientado NE-SW.
Gr. Natividade	GrNat	Metassedimentos variados: quartzitos, metarenitos, conglomerados, metapelitos, carbonatos.
St. Peraluminosa	SP	Muscovita-biotita metagranodiorito com ocorrência subordinada de metatonalitos e metamonzogranitos
St. Peraluminosa - tonalítica	SPT	Stocks de metatonalito leucocrático com menos de 10% de biotita.
Suíte Quartzo-Diorítica	SQD	Composta de dois grandes corpos e algumas intrusões de menor porte de quartzo-metadiorito, metagabros e subordinadamente metatonalitos
St. Ribeirão Itaboca	RIS	Biotitas metatonalitos e gnaisses tonalíticos. Essa unidade é caracterizada por apresentar feições de migmatítico gnáissico.
St. Ribeirão das Areias	RAC	Metatonalitos mesocráticos e leucocráticos e metatrondhjemitos com ocorrências subordinada de granodioritos. Ocorrem como batólitos com formato sigmoidal orientados NNW-SSE a NNE-SSW
St. Ribeirão das Areias - corpos intrusivos subordinados	RAC-IB	Metagranodiorito de granulação média a grosseira e intrusões monzograníticas
St. Serra Boqueirão	SSB	Caracterizada por um hornblenda metatonalito de granulação média a grossa, mesocrático, com texturas inequigranulares e foliação
Fm. Morro do Carneiro	FMC	Formada por metassedimentos terrígenos e químicos, composta por filito sericitico carbonoso, quartzito, metassilexito, xisto carbonoso, formação ferrífera bandada, metaconglomerado e, subordinadamente, rochas metavulcânicas félsica (muscovita-plagioclásio-quartzo xisto), intermediária (clorita-anfibólio xisto feldspático) e máfica (anfibólio xisto e anfibolito)
Fm. Córrego do Paíol	FCP	Composta predominantemente de metabasaltos com raras ocorrências de rochas ultramáficas. É descrito dois tipos de magmas fonte, os magmas toleíticos ricos em Fe e subordinados basaltos komatiitos ricos em Mg.

3. Materiais e Métodos

Os dados utilizados no presente trabalho são de caráter público, e estão disponíveis na plataforma *GeoBank* do Serviço Geológico Brasileiro, com exceção dos dados adquiridos durante a campanha de campo. A base de dados utilizada para geração dos mapas litológicos preditivos foi composta por quatro tipos de dados: Cartografia Geológica, dados aeromagnetométricos, dados aeroradiométricos e dados altimétricos. A partir destes dados alguns produtos foram gerados com o objetivo de gerar contraste entre valores de propriedade física de diferentes tipos de rocha.

3.1 Dados Litológicos

A confecção do mapa litológico utilizado durante as próximas etapas tem como base a Folha Dianópolis (Saboia *et al.*, 2019), Folha Arraias (Abdallah & Meneghini, 2017), o mapa geológico elaborado por Saboia (2021) por diferenciar em um maior

detalhe as unidades pertencentes ao embasamento granítico-gnaissico, e por fim o mapa geológico do estado do Tocantins elaborado pelo Serviço Geológico Brasileiro (Ribeiro *et al.*, 2022) pois nele novos corpos pertencentes a Suíte Gameleira foram inseridos.

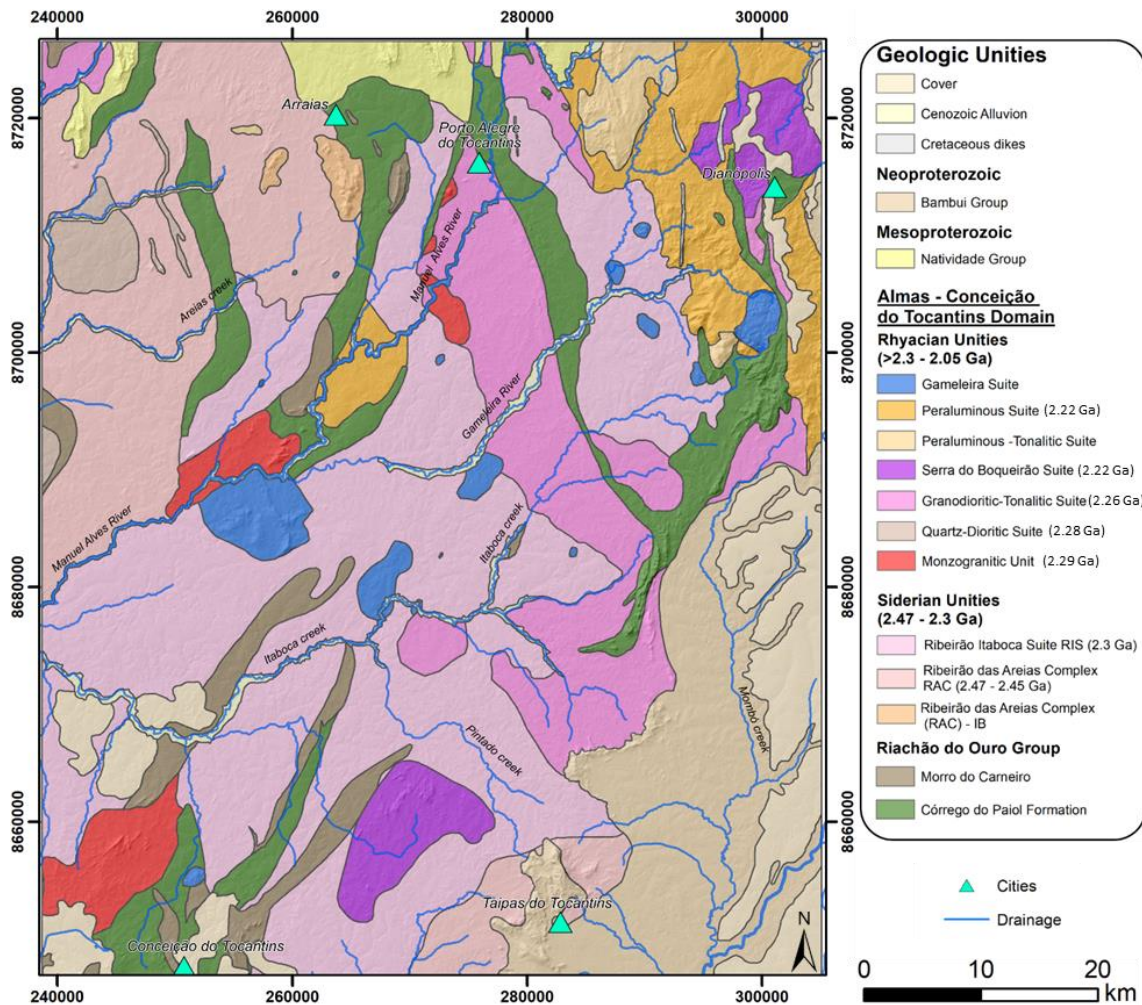


Fig. 1. Mapa litológico extrapolado da área de estudo, modificado de Saboia, 2021; Saboia *et al.*, 2019; Abdalah *et al.*, 2015; Ribeiro *et al.*, 2022.

Foi realizada uma campanha de campo com o intuito de reconhecer a forma de ocorrência e a mineralogia das rochas da Suíte Gameleira. Foi gerado com o auxílio dos dados aerogeofísicos e imagem de satélite um conjunto de dados pré-campo com base nos dados aerogeofísicos dando ênfase em regiões anômalas quanto a teores radiométricos (baixa resposta nos três canais) e altos valores de intensidade magnética. Essa etapa tem grande importância, uma vez que os corpos delimitados por mapeamentos previamente realizados não trazem informações de pontos de campo ou a relação dos litotipos de cada corpo. A vista disso foram visitados e amostrados alguns dos corpos já mapeados pelas folhas geológicas que abrangem a área de estudo, que possibilitaram a

discriminação de alguns litotipos máficos, como por exemplo leucogabros, hornblenda gabro, gabro-norito, e ultramáficos como peridotito, dunitos e serpentinitos.

3.2 Dados Aerogeofísicos

Os produtos aerogeofísicos, aeromagnetométricos e aeroradiométricos, utilizados no projeto foram gerados a partir da base de dados pertencente ao aerolevanteamento do Projeto Aerogeofísico Tocantins (CPRM, 2006), com linhas de voo e de controle espaçadas de 0,5 km e 10 km, orientadas nas direções N-S e E-W, respectivamente, velocidade média de voo de 270 km/h, e altura de voo fixada em 100 metros sobre o terreno.

3.2.1 Dados Aeroradiométricos

A radiometria, também conhecida como gamaespectometria, é um método potencial com o objetivo de detecção e quantificação de radiação ionizante, a partir da emissão natural de raios gama (γ) por decaimento radioativo de isótopos (Wilford & Minty, 2006). Mais de 50 elementos naturalmente se comportam de maneira radioativa, no entanto a radiação terrestre é dominada pelo produto da emissão de radioatividade de três elementos: potássio (K), tório (Th) e urânio (U), onde somente esses três radioelementos emitem raios gama suficientes para serem detectados e medidos em levantamentos aéreos (Dentith & Mudge, 2014; Minty, 1997).

A detecção de radiação por meio de aerolevanteamentos tem um alcance de aproximadamente 30cm, e é amplamente utilizada em mapeamentos geológicos. Alguns fatores podem influenciar os valores das medidas de radiação, dentre eles: vegetação; grande variação topográfica; efeitos direcionais; e as características do aerolevanteamento (Ribeiro, 2014).

3.2.1.1 Processamento e Realce dos Dados Aeroradiométricos

O processamento dos dados gamaespectométricos teve início na análise dos dados, através de parâmetros estatísticos como a média, mediana, e valores máximos e mínimos para os canais de Contagem Total (cps), K (%), eTh (ppm) e eU (ppm) (Fig. 2 – A, B, C e D, respectivamente). Com a análise inicial identificou-se a presença de teores negativos no banco de dados, o que representa uma inconsistência física para uma fonte radioativa, para a correção desses valores foi aplicada uma normalização (Equação 1) com o intuito de não mais haver valores negativos e zeros, uma vez que razões seriam realizadas posteriormente (Mendes, 2023 no prelo).

$$X_{normalized} = \frac{(X - X_{min})}{(X_{max} - X_{min})} + 0.001 \quad \text{Equação 1}$$

Os dados então foram interpolados para células de 125 metros, um quarto do espaçamento das linhas de voo, com o uso do interpolador mínima curvatura (Briggs (1974) e Swain (1976)) através do *software Oasis Montaj*. A partir, destes dados foram realizadas outras razões, normalizações que visaram obter assinaturas que possibilitassem um melhor contraste das unidades.

Dentre elas o Fator F (1978 apud Gnojek e Prichystal, 1985) evidenciando o enriquecimento de K e urânio em relação ao tório (Fig. 2 – E) através da Equação 2, regiões com baixos valores podem possuir uma concentração de Tório equivalente alta, comparado aos arredores, e da mesma forma localidades em que se espera um valor menor e foram tidos como mais alto podem vir a terem uma relação com feições hidrotermais.

$$F = \frac{K*U}{Th} = \frac{U}{\frac{Th}{K}} = \frac{K}{\frac{Th}{U}} \quad \text{Equação 2}$$

Outras duas razões em função do tório foram realizadas com o intuito de reduzir as variações primárias causadas por variação de litologia superficial, tipos de solos, e condições ambientais como umidade e vegetação, para uma melhor compreensão de processos secundários como acúmulo de determinados minerais e segregação de fluídos por resultado de frentes de alteração hidrotermal. Para Saunders *et al.*, (1987), os valores ideais de urânio e potássio podem ser definidos para cada amostra tendo como base os teores de tório (Equação 3 e 4), uma vez que os efeitos litológicos e ambientais que afetam a concentração de tório afetam também os teores de potássio e urânio de modo previsível (Saunders *et al.*, 1993).

$$K_i = \left(\frac{\text{média de } K_s}{\text{média de } Th_s} \right) * Th_s \quad \text{Equação 3}$$

$$U_i = \left(\frac{\text{média de } U_s}{\text{média de } Th_s} \right) * Th_s \quad \text{Equação 4}$$

Na Equação 3 e 4, as variáveis Ks, Ths e Us correspondem aos valores de potássio, tório e urânio medidos em cada ponto, já Ki e Ui correspondem aos valores ideais de potássio e urânio, respectivamente, normalizados em função do tório. Para o cálculo dos desvios entre os valores reais com base nos valores ideais, utiliza-se as equações 5 e 6:

$$K_d = \frac{(K_s - K_i)}{K_i} \quad \text{Equação 5}$$

$$U_d = \frac{(U_s - U_i)}{U_i} \quad \text{Equação 6}$$

Em que K_d , (Fig. 2 – F) e U_d correspondem aos valores de concentração anômalas para o potássio e o urânio (Saunders, 1987, 1993; Pires, 1995).

3.2.2 Dados Aeromagnetométricos

A magnetometria aérea é muito utilizada para a diferenciação de domínios litológicos, dado que diferentes rochas vão possuir diferentes assinaturas de susceptibilidade magnética. Na área de estudos temos um conjunto de rochas com diferentes valores de susceptibilidade magnética, são esses metassedimentos a rochas intrusivas ácidas a ultrabásicas.

A área de estudo marcada pela presença de magnetização remanente por parte de suas rochas com conteúdo máfico, a Suíte Gameleira, uma maneira qualitativa de avaliar se há ou não remanência é pela identificação de dipolos invertidos (Austin *et al.*, 2019) a unidade alvo da pesquisa tem suas ocorrências por vezes caracterizada por assinaturas dipolares invertidas. A magnetização remanescente em rochas máficas é resultado de um ou mais dos seguintes eventos distintos: 1) Cristalização e subsequente resfriamento através da temperatura Curie; 2) Ex-solução de titanomagnetita a partir de magnetita e ilmenita em torno de 600 °C; 3) Aquecimento e reajuste paleomagnético durante o metamorfismo e/ou resfriamento após a exumação de rochas da crosta média (Austin *et al.*, 2019).

3.2.2.1 Processamento e Realce dos Dados Aeromagnetométricos

Para a análise magnética da área de estudo os dados foram interpolados para células de 125 metros, um quarto do espaçamento das linhas de voo, com o uso do interpolador bi-direcional através do *software Oasis Montaj*. O dado utilizado para a geração do produto de Intensidade Magnética Total (TMI) foi o dado de intensidade do campo magnético fornecido na base de dados da CPRM, 2006, já com o IGRF (*International Geomagnetic Reference Field*) removido.

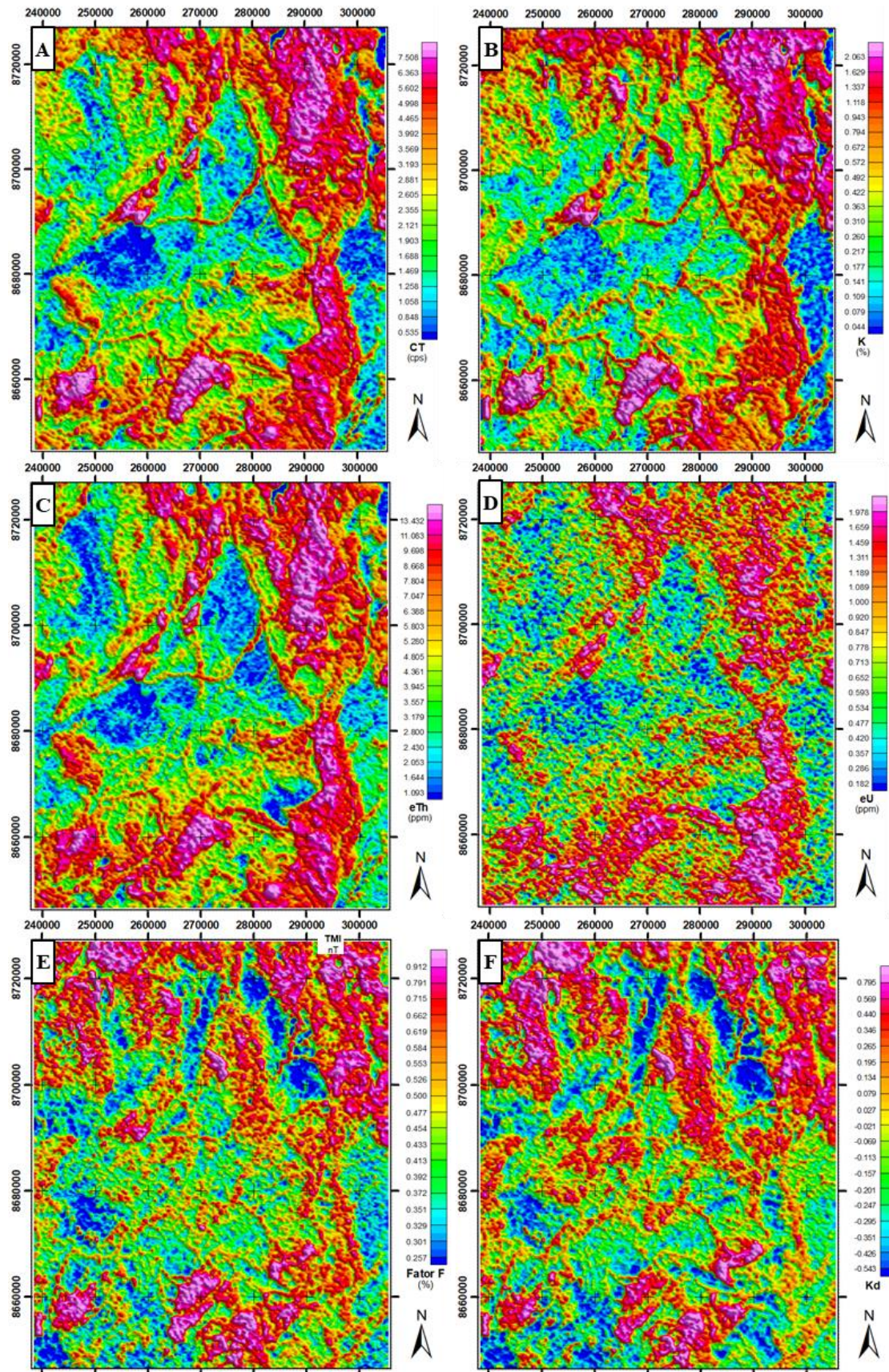


Fig. 2. Produtos aeroradiométricos gerados para a área de estudo. A – Contagem Total; B – Canal do Potássio; C – Canal do Tório; D – Canal do Urânio; E – Fator F; F - Potássio Anômalo.

Calculou-se filtros, realces e confeccionou-se modelos de inversão 3D a partir do TMI (Fig. 3 – A) uma vez que a técnica de Redução ao Polo (Baranov, 1957; Bhattacharya, 1965) é instável em baixas latitudes e tem a tendência de não apresentar bons resultados em locais com forte remanência magnética, gerando um viés (Nabighian *et al.*, 2005). Características comuns na área de estudo, uma vez que a mesma está inserida na Anomalia Magnética do Atlântico Sul (Chapman & Bartels, 1940).

Os principais produtos calculados a partir do TMI foram a Amplitude do Sinal Analítico (ASA) (Fig. 3 – B) e o Gradiente Horizontal Total (GHT) (Fig. 3- C), muito usados para a compreensão da geometria da fonte magnética e detecção de limites de corpos (Nabighian 1972; Blakely, 1996; Milligan & Gunn, 1997; Dentith & Mudge, 2014). Os filtros referentes a primeira derivada vertical (Nabighian, 1984; Dentith & Mudge, 2014) e a derivada Tilt (Miller & Singh, 1994; Verduzco, 2004) foram utilizados para a extração dos lineamentos magnéticos e correlação com os corpos da Suíte Gameleira indicados pelo algoritmo, visando uma melhor compreensão de áreas de pouca assertividade.

Foi gerado um índice máfico (Pires & Moraes, 2006) para a área de estudo com o intuito de individualizar corpos máficos e ultramáficos. O produto relaciona os dados magnéticos e radiométricos por meio da razão entre a Amplitude do Sinal Analítico e o produto dos três canais de radioelementos (K, eTh e eU) através da equação 7 foi gerado (Fig. 3 - E).

$$IM = \frac{ASA}{K * eTh * eU} \quad \text{Equação 7}$$

Um último produto magnético gerado com o intuito de ser um dos features para a aplicação do Random Forest foi um Slice horizontal de -100m de profundidade do retirado da Inversão dos Vetores de Magnetização (MVI) (Fig. 3 - D), como proposto por Costa (2019) visando filtrar fontes mais rasas e imagear somente as mais profundas como espera-se que os corpos da Suíte Gameleira se comportem.

Os modelos de inversão MVI foram gerados mediante o *software Oasis Montaj*, em seu menu VOXI, o algoritmo leva em consideração os efeitos de magnetização remanente e a posição em baixas latitudes (Ellis *et al.*, 2012; Macleod & Ellis, 2013; Barbosa, 2013; Johnson & Aisengart, 2014; Aisengart, 2015). De maneira que os modelos de inversão foram utilizados como uma maneira de verificação remota dos mapas

preditivos, uma vez que apresentassem uma anomalia em profundidade e/ou uma perturbação característica dos vetores de magnetização.

4. Implementação do Random Forest

4.1 Pré-processamento e elaboração da base de dados

No presente trabalho foi elaborado um mapa litológico “extrapolado” de maneira que os mapas geológicos (Abdalah *et al.*, 2015; Saboia *et al.*, 2019; Saboia, 2021; Ribeiro *et al.*, 2022), com escalas semelhantes, foram analisados e quando necessário, ajustados para melhor representação da área. O produto final desta etapa foi o mapa litológico da área de estudo (Fig. 1), totalizando 17 classes litológicas, extraídas do mapa geológico, que corresponderam aos rótulos para a aprendizagem do modelo.

Uma vez que todos os produtos aerogeofísicos foram confeccionados, foi realizada reamostragem para células de 250 metros, gerando uma malha regular com tamanho 66 km x 80 km (5.293 Km²), aproximadamente, e com espaçamento de 250m. Essa base de dados foi exportada para *Software ArcMap* para que valores de altimetria (Fig. 3 – F) do satélite ALOS sensor Palsar fossem extraídos para cada ponto da malha. Procedimento semelhante foi realizado no mapa litológico da área de estudo, onde aplicou-se uma manipulação para que em cada ponto da malha gerada fosse atribuída uma respectiva classe litológica (rótulo) correspondente à posição do ponto no mapa geológico.

Dessa maneira o banco de dados utilizado conta com total de 84.535 instâncias, com 16 colunas de variáveis, contando com as coordenadas espaciais (longitude e latitude em UTM) e 17 classes. As variáveis com coeficiente de correlação de Pearson >0.8 quando associadas a outras variáveis foram eliminadas, de maneira que se utilizou somente 9 variáveis para o treino do algoritmo e realização de previsões para geração do modelo final.

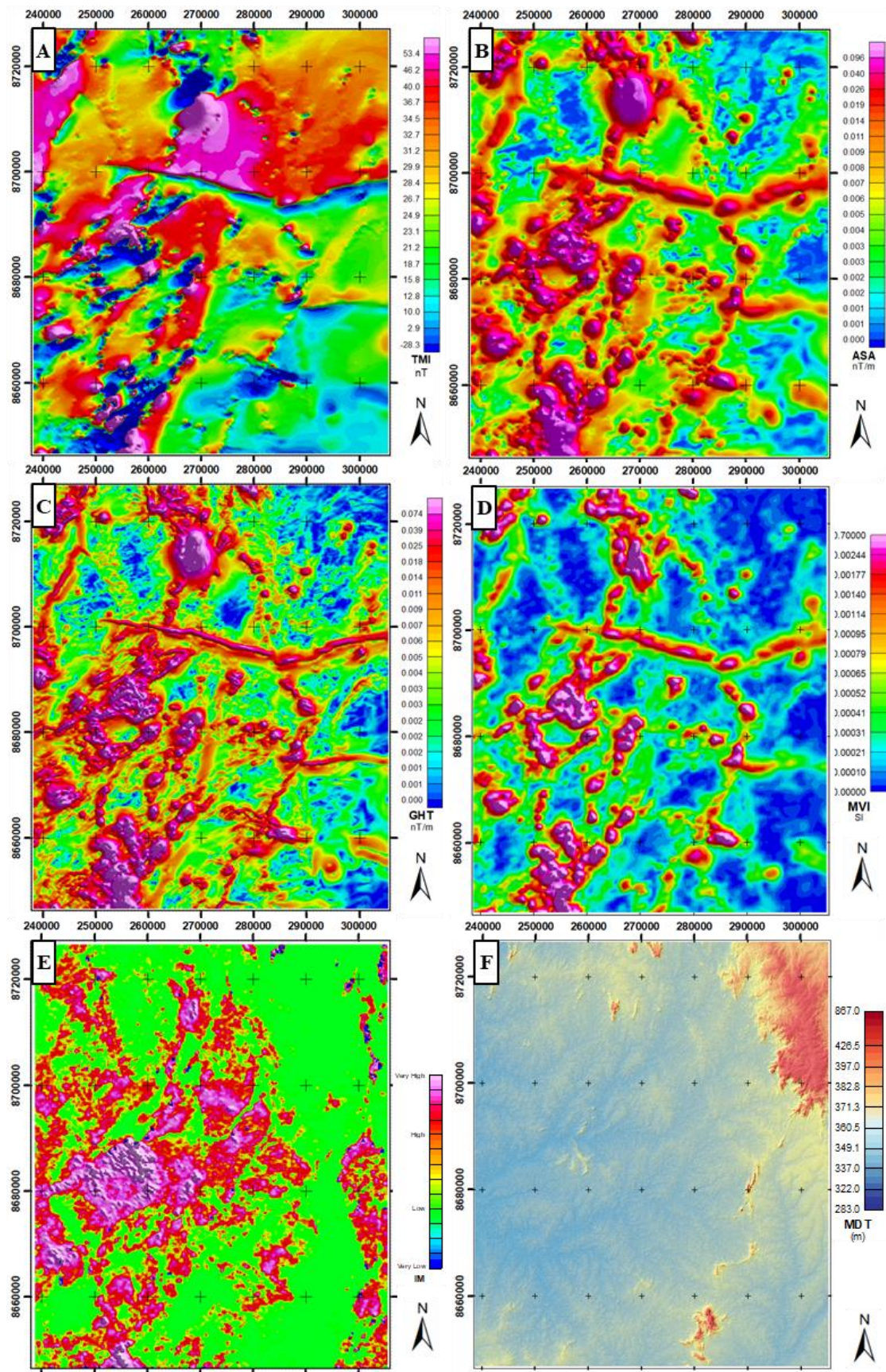


Fig. 3. Produtos aeromagnéticos gerados para a área de estudo. A – Intensidade Magnética Total; B – Amplitude do Sinal Analítico; C – Gradiente Horizontal Total; D – *Magnetic Vectors Inversion* – slice de -100 metros de profundidade; E – Índice Máfico; F – Altimetria.

4.2 Random Forest

O *Random Forest* é um algoritmo de aprendizado de máquinas supervisionado do tipo *ensemble* (conjunto), onde é gerada uma combinação de árvores de classificação e regressão no qual a cada uma é fornecido um vetor amostrado independentemente, de forma aleatória, e apresenta a mesma distribuição (Breiman, 2001). A aleatoriedade é inserida na implementação do algoritmo pelo crescimento de árvores para um subconjunto aleatório, a partir de um número pré-definido de variáveis para dividir cada nó das árvores de decisão e por um processo denominado *bootstrap aggregation* conhecido como *bagging* (Breiman, 1996). Os parâmetros de entrada no classificador de *Random Forest* são mínimos e incluem apenas o número de variáveis para cada árvore e o número de árvores para criar (Harris, 2015). O número de variáveis foi definido para a raiz quadrada do número total de variáveis e o número de árvores foi definido para cada modelo através da experimentação.

Uma das maneiras de se analisar a aderência dos dados de treino ao algoritmo é a validação cruzada, onde essa abordagem divide as instâncias de treino em diversas dobras (*folds*) e utiliza-se cada uma dessas *folds* por vez para realizar um treinamento e predição do remanescente. Esse processo é repetido ‘n’ vezes, até que todas as *folds* tenham sido treinadas como um conjunto de validação, em cada modelo é avaliada performance e ao fim uma média com os ‘n’ valores das métricas calculadas é fornecido (Costa *et al.*, 2019, Burkov, 2019).

A aplicação do algoritmo Random Forest (Breiman, 2001), foi para a realização da classificação litológica com o uso de somente três tipos de dados como variáveis: magnéticos, radiométricos e altimétricos e seus subprodutos.

4.2.1 Dados de treino e classificação

A estratégia de treinamento foi de fixar um número de 100 amostras aleatórias para cada classe, como uma estratégia de minimizar o viés gerado por um conjunto de amostras desbalanceado (Provost & Fawcett 1997; Japkowicz & Stephen 2002)., com exceção dos diques Cenozoicos, por possuírem número inferior a 100, para essa classe foi treinado 70% de seu total, totalizando uma base de treino de 1659 (Fig. 4). As amostras da base de treino correspondem a 1,96% da base de dados, valor próximo aos valores utilizados por Cracknell *et al.*, (2014a), Kuhn *et al.*, (2018) e Costa *et al.*, (2019), que utilizaram 2.6%, 1.6% e 2.5%, respectivamente.

Para cada experimento foi realizada a seguinte rotina: (i) cálculo do melhor número de árvores para o conjunto de variáveis utilizadas, a partir da validação cruzada com 10 *folds*, e análise do desempenho do modelo; (ii) análise do peso de cada variável dado pelo modelo; (iii) realização da predição e análise de performance; (iv) elaboração de uma matriz de confusão;

A implementação do algoritmo será apresentada para sete testes realizados, com intuito de analisar no que a variação de entradas (*features*) modificou o modelo, e quais seriam as melhores combinações para a prospecção de corpos da Suíte Gameleira. Para a avaliação do modelo e predição foram utilizadas três escalas de métricas como referência sendo elas a Classificação de Acurácia, F1-score e a Precisão, por se tratar de um conjunto de dados desbalanceado a classificação de acurácia nem sempre é a melhor maneira de representar o desempenho do modelo para cada classe (Galar *et al.*, 2011).

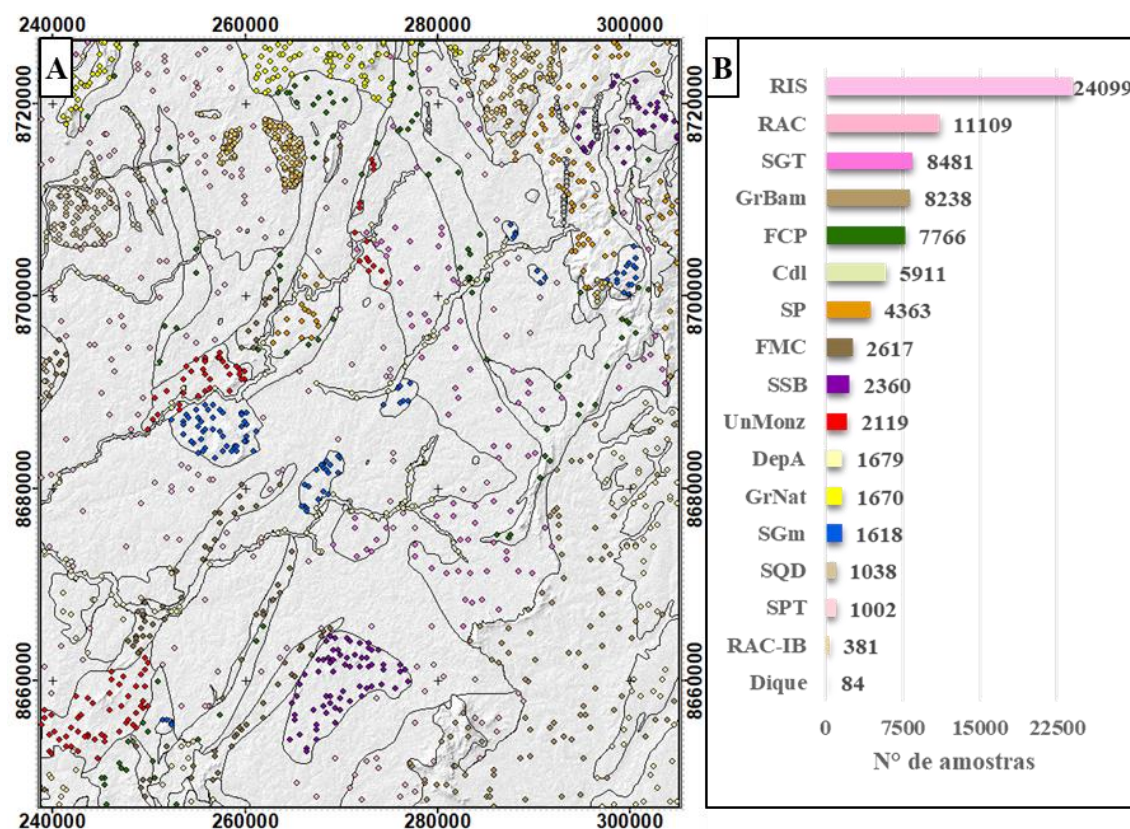


Fig. 4. A - Localização dos pontos de treinamento utilizado nos experimentos para implementação do Random Forest. A base de dados de treino é composta por 1659 pontos, sendo 100 amostras aleatórias de cada classe, e 59 pontos da classe dos diques cenozoicos; B – Diagrama indicando a proporção do número de amostras para cada uma das 17 classes litológicas. O eixo Y corresponde as seguintes unidades: Coberturas detrito-lateríticas (Cdl), Depósitos aluvionares (DepA), Diques (Dique), Gr. Bambui (GrBam), St. Gameleira (SGm), St. Granodiorítica aTonalítica (SGT), Un. Monzogranítica (UnMonz), Gr. Natividade (GrNat), St. Peraluminosa (SP), St. Peraluminosa-tonalítica (SPT), Suíte Quartzo-Diorítica (SQD), St. Ribeirão Itaboca (RIS), St. Ribeirão das Areias (RAC), St. Ribeirão das Areias (RAC-IB), St. Serra Boqueirão (SSB), Fm. Morro do Carneiro (FMC) e Fm. Córrego do Paiol (FCP).

5. Resultados

5.1 Dados de campo

Durante a etapa campo, 7 ocorrências da Suíte Gameleira foram visitadas (Fig. 5), onde caminhamentos foram realizados para encontrar rochas preservadas, uma vez que rochas máficas e ultramáficas alteram-se com facilidade e os processos intempericos na área de estudo são pervasivos.

As ocorrências das rochas da unidade em grande maioria são blocos *in-situ* ou rolados, apresentando comportamento aparentemente isotrópico, com presença de estruturas (bandamentos composicionais, foliação, estiramento mineral) quando próximas a zonas de cisalhamento locais. Durante a etapa de campo, foram descritas rochas que correspondem aos litotipos hornblenda gabro (Fig. 5 – A, G), leuco gabro (Fig. 5 – E, F), gabro norito (Fig. 5 – D), olivina gabro norito (Fig. 5 – B, H), dunitos serpentinizados (Fig. 5 – C).

A ocorrência com maior variação de litotipos, e onde as rochas da Suíte Gameleira se encontram claramente tectonizadas, é a IBG (Correia & Sá, 1980, CPRM, 2000, Mendes *et al.*, 2023). As rochas dessa intrusão correspondem majoritariamente a máficas com presença de ultramáficas condicionadas por estruturas tectônicas. A deformação é visível em amostras macroscópicas pelo estiramento mineral (Fig. 5 –B) e por foliação protomilonítica, localmente encontrada no afloramento onde foi coletada amostra (Fig. 5 – C) com impregnação de malaquita em planos de fratura.

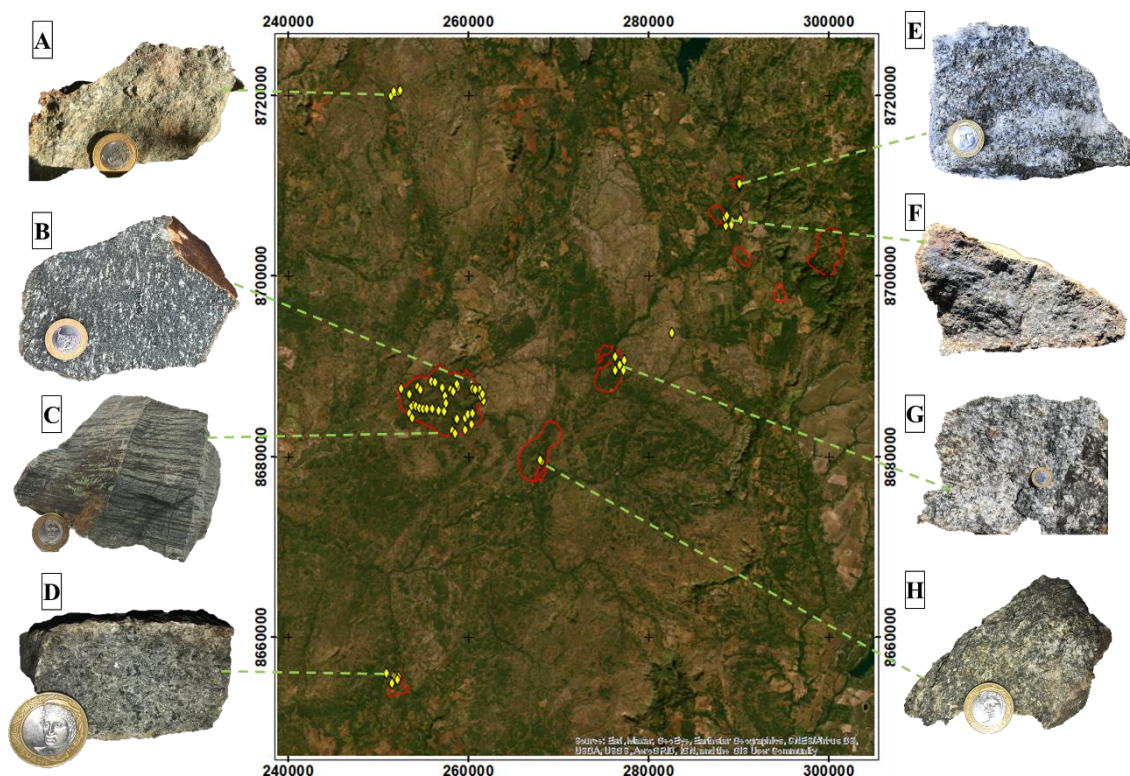


Fig. 5. Pontos de visita  o em campo e litotipos descritos. A e G – Hornblenda Gabro; B e H – Olivina Gabro Norito; C -Dunito Serpentinizados; D – Gabro Norito; E e F - Leuco Gabros. Em vermelho o contorno das unidades cartografadas no mapa litol gico.

5.2 An lise explorat ria dos dados

Foi realizada an lise para a compreens o dos dados que consistiu na averigua  o dos tipos de distribui  o de cada par metro f sico, analisando individualmente a distribui  o de cada um. Pelo fato de a base de dados ser extensa e testes como o de Shapiro-Wilk (Shapiro & Wilk, 1965; Royston, 1982) n o trazerem uma boa resposta para *datasets* maiores que 5000 amostras. Neste trabalho avaliou se as distribui  es eram gaussianas por meio da utiliza  o de gr ficos *QQPlot* e gr ficos do tipo *Box Plot* (Oppong & Agbedra, 2016). Os gr ficos *QQPlot* (Fig. 6) indicam que nenhum dos par metros apresenta distribui  o normal para o conjunto de amostras, de maneira que estes dados n o s o param tricos, por isso a m dia e o desvio padr o n o s o estimativas representativas para a an lise do centro e distribui  o dos dados.

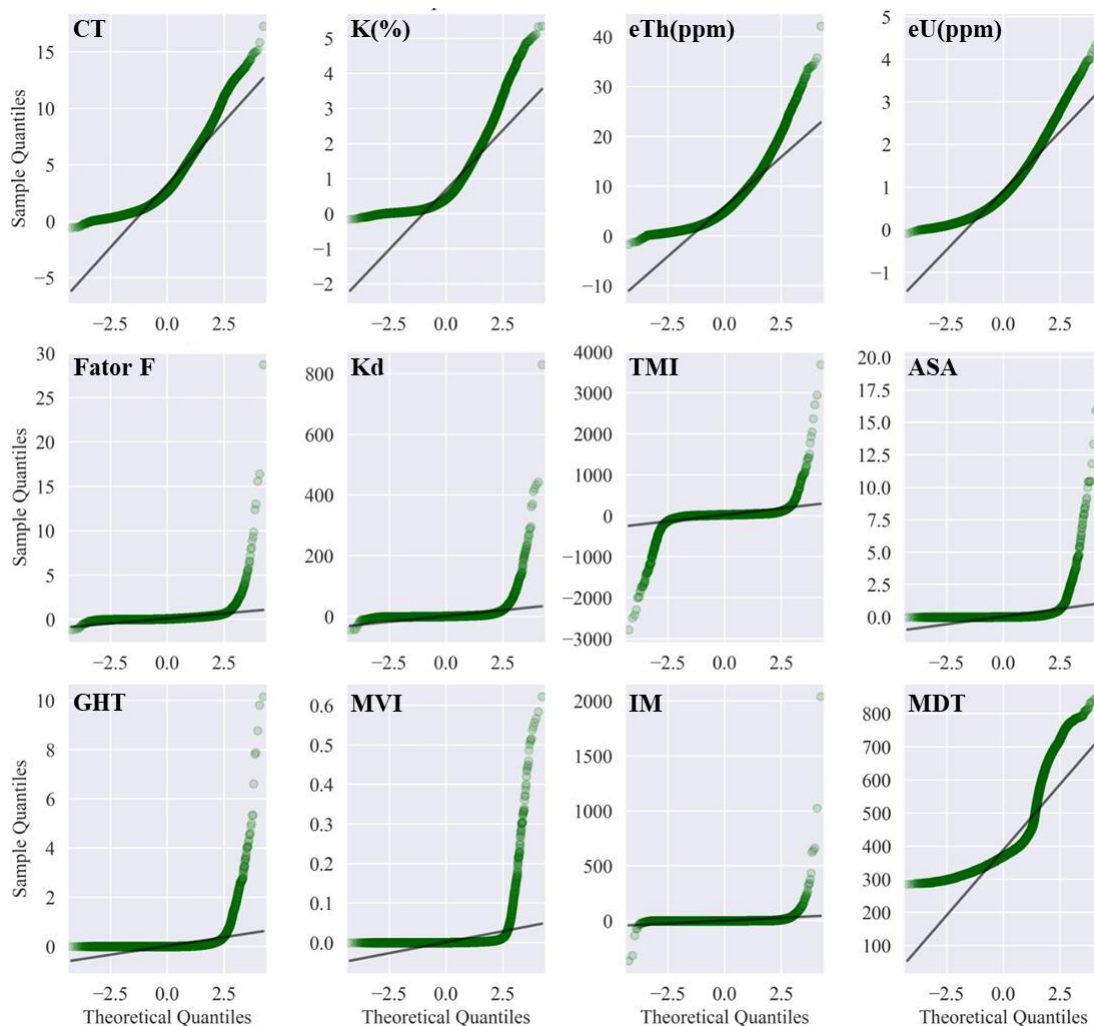


Fig. 6. Diagrama Quantil-Quantil (QQ-plot) para os parâmetros físicos utilizados. Em verde a distribuição dos pontos, a linha preta indica o comportamento esperado para uma distribuição normal.

A base de dados foi subdividida para cada unidade e teve seus parâmetros K, eTh, eU e ASA interpretados com o uso do diagrama *Box & Whiskers*, tomando como estimativa a mediana. No caso, os parâmetros analisados apresentam um comportamento bem difuso nas unidades, no caso do diagrama do potássio (Fig. 7 – A) a unidade Serra do Boqueirão (SSB) e a St. Monzogranítica (UnMonz) apresentam uma alta variabilidade com poucos outliers em quanto as unidades Cobertura (Cdl) e da St. Gameleira (SGm), por exemplo, apresentam uma baixa variabilidade e são marcadas pela presença de *outliers*.

Há uma presença significativa de *outliers* em algumas unidades, esses *outliers* caracterizam uma distribuição assimétrica com cauda para a direita, e podem ser resultantes de pontos amostrados que podem pertencer a outra unidade, subgrupos que

correspondem a uma variação faciologica dentro da própria unidade ou mesmo ocorrências anômalas de algum parâmetro. Todas essas relações podem ser correlacionadas a questão de detalhamento e escala do mapeamento e do resultado.

O diagrama do K (Fig. 7 – A) quando analisado em ordem decrescente de concentração por unidade, indica mudança relativa no valor da mediana a partir do Grupo Natividade (GrNat), onde as unidades a esquerda são todas intrusivas ácidas, com exceção dos diques cretáceos que são mapeados dentro do domínio dessas intrusivas e raramente afloram, essa variação ocorre justamente quando há uma transição da química da rocha. Neste contexto a direita do Grupo Natividade temos três unidades intrusivas ácidas e uma intrusiva intermediária, no entanto, essas três unidades apresentam número elevado de *outliers*, e possuem ampla cobertura no mapa litológico. A distribuição dos radioelementos eTh e eU (Fig. 7 – B e C) indicam assimetria e grande concentração de *outliers*, no entanto não foi possível de maneira qualitativa realizar interpretações que possibilitassem o agrupamento de unidades por seus valores. Para esses radioelementos ressalta-se que a Suíte Gameleira apresenta o menor teor, tendo como referência a mediana, quando comparada as demais unidades o que está de acordo com o padrão esperado (Dickson e Scott, 1997). No diagrama referente aos valores do produto ASA, em nT/m, a unidade que apresenta o mais alto valor de intensidade usando como referência a mediana é justamente a Suíte Gameleira, indicando assinatura característica da unidade, ao passo que as demais unidades que possuem valores elevados são a Formação Morro do Carneiro, Grupo Natividade e Formação Córrego do Paiol.

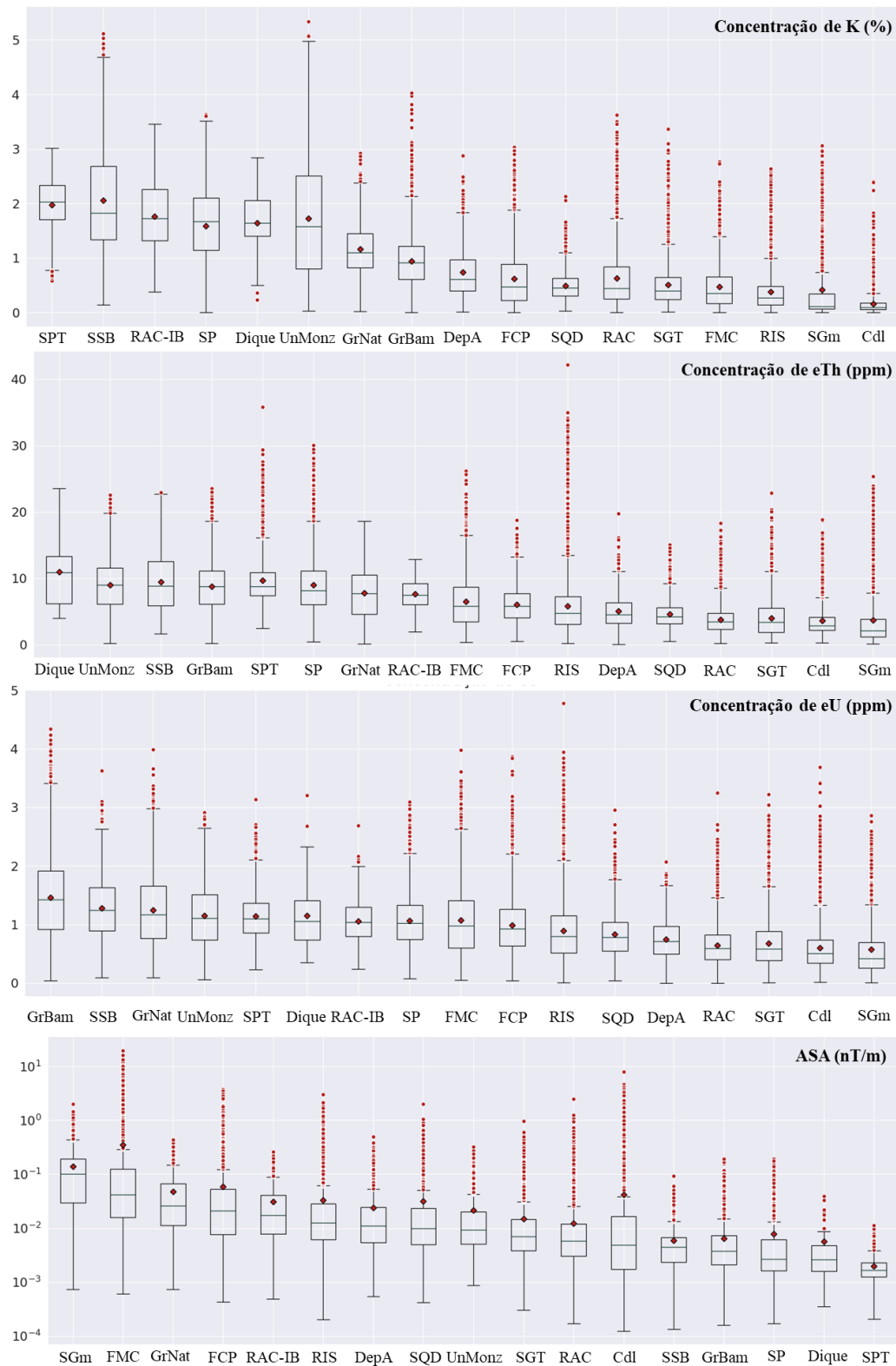


Fig. 7. Diagramas do tipo Box Plot, elaborado para caracterização e visualização da assinatura de cada unidade, classe, com relação aos parâmetros físicos K, eTh, eU, ASA, respectivamente, A, B, C e D. O eixo X corresponde as seguintes unidades: Coberturas detrito-lateríticas (Cdl), Depósitos aluvionares (DepA), Diques (Dique), Gr. Bambui (GrBam), St. Gameleira (SGm), St. Granodiorítica aTonalítica (SGT), Un. Monzogranítica (UnMonz), Gr. Natividade (GrNat), St. Peraluminosa (SP), St. Peraluminosa-tonalítica (SPT), Suíte Quartzo-Diorítica (SQD), St. Ribeirão Itaboca (RIS), St. Ribeirão das Areias (RAC), St. Ribeirão das Areias (RAC-IB), St. Serra Boqueirão (SSB), Fm. Morro do Carneiro (FMC) e Fm. Córrego do Paiol (FCP).

5.3 Mapas de classificação utilizando Random Forest

A Fig. 8 corresponde aos resultados do treinamento do algoritmo pela informação dos três canais radiométricos, altimetria e a variações do parâmetro magnético. A motivação para esses testes frente aos parâmetros magnéticos vem da estatística descritiva, que indica que a assinatura magnética da Suíte Gameleira é o parâmetro que mais se distingue quando comparado a outras unidades. O resultado da classificação de acurácia, F1-score e Precisão dos modelos e predições para os resultados estão na Tabela 2.

Tabela 2: Informação de parâmetros utilizados na realização de testes e suas métricas resultantes.

Teste	Árvores	Parâmetros	CA		F1-Score		Precision	
			Modelo	Predição	Modelo	Predição	Modelo	Predição
A-01	500	CMA, Knorm, Thnorm, Unorm, MDT	0.643	0.388	0.632	0.403	0.638	0.502
A-02	1000	CMA, Knorm, Thnorm, Unorm, MDT, X	0.862	0.569	0.86	0.576	0.861	0.667
A-03	1000	ASA, Knorm, Thnorm, Unorm, MDT	0.591	0.358	0.579	0.363	0.586	0.46
A-05	750	MVI, Knorm, Thnorm, Unorm, MDT	0.605	0.359	0.593	0.363	0.601	0.459
A-07	1000	CMA, ASA, MVI, Knorm, Thnorm, Unorm, MDT	0.744	0.448	0.737	0.458	0.746	0.558
A-09	750	CMA, ASA, MVI, Knorm, Thnorm, Unorm, Imnorm, Kdnorm, MDT	0.741	0.453	0.734	0.465	0.743	0.558
A-10	751	CMA, ASA, MVI, Knorm, Thnorm, Unorm, Imnorm, Kdnorm, MDT, X	0.888	0.594	0.885	0.603	0.887	0.689

Nota-se que a reconstrução da Suíte Gameleira, foi distinta para cada teste de maneira que o teste A-01 forneceu um formato mais aproximado, do mapa litológico de referência, para os corpos localizados a nordeste, enquanto os testes A-03 e A-05 conseguiram realizar uma melhor reconstrução dos corpos na porção central do mapa. Quando utilizado as três *features* magnéticas, obteve-se melhor qualidade de classificação e uma possível melhor reconstrução dos corpos, com exceção do corpo no extremo nordeste pela presença de uma fonte magnética de direção N-S, bem marcada no produto do MVI (Fig. 3 – D) pode ter funcionado como um viés na classificação.

O teste A-09 resultou em um dos melhores resultados de classificação, bem como numa reconstrução mais próxima dos domínios litológicos, predispostos, foi resultante da integração das 9 variáveis (TMI, ASA, MVI, K, eTh, eU, Kd, Alos e IM), este apresentou uma classificação de acurácia, F1-score e Precisão com valores de 0.453, 0.465 e 0.558, respectivamente (Tabela 2).

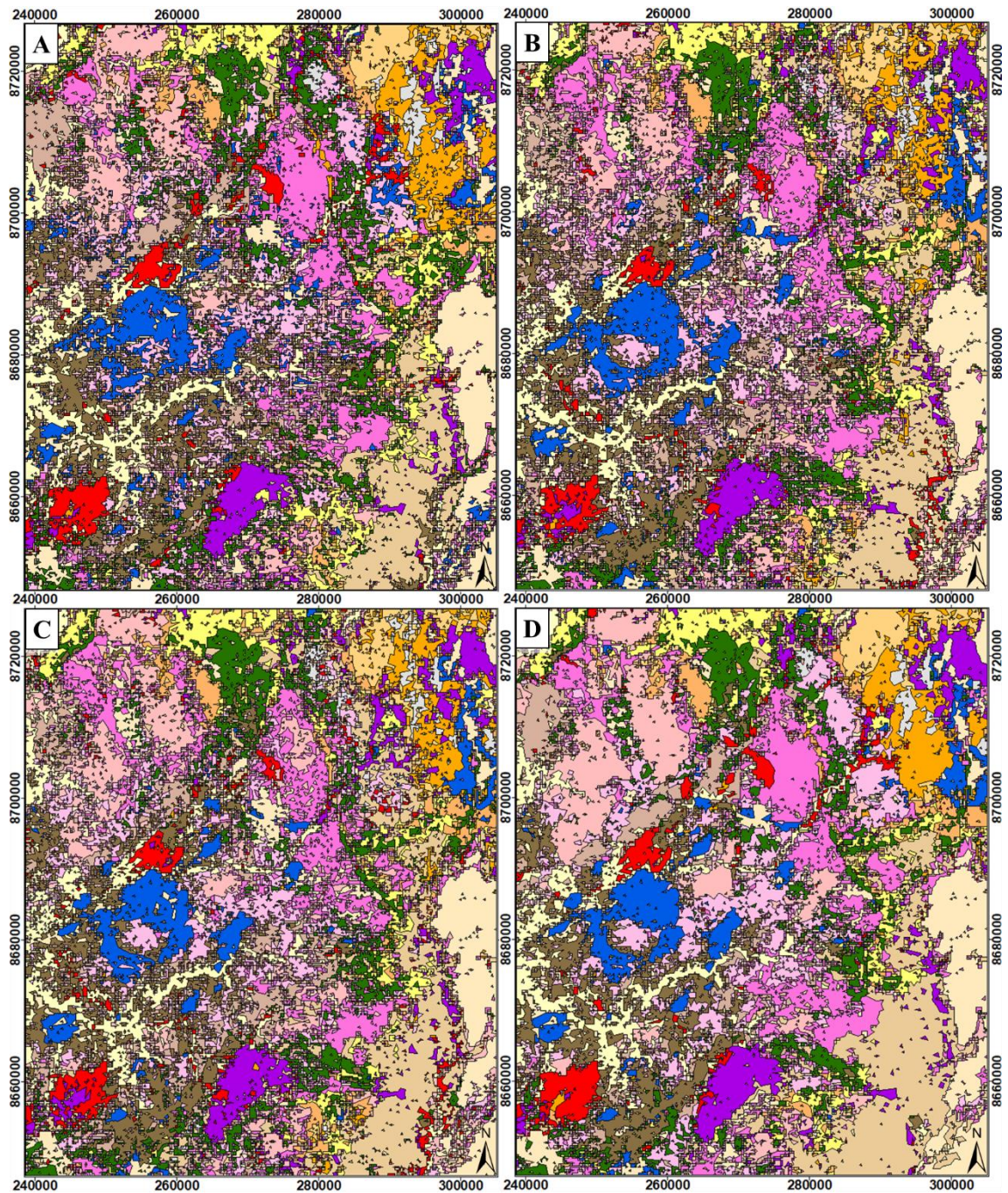


Fig. 8. Mapas resultantes da classificação litológica realizada pelo algoritmo Random Forest com base nos três canais radiométricos (K, eTh e eU), altimetria (Alos) diferentes entradas de dados magnéticos. A – Teste A-01 (TMI, K, eTh, eU e MDT); B - Teste A-03 (ASA, K, eTh, eU e MDT); C - Teste A-05 (MVI, K, eTh, eU e MDT); D - Teste A-07 (TMI, ASA, MVI, K, eTh, eU e MDT).

No mapa preditivo gerado pelo teste A-09 (Fig. 9 –A) algumas unidades tiveram seu tamanho reduzido enquanto outras o oposto, como é o caso da unidade Ribeirão de Itaboca (RIS) em que a predição reconstruiu somente 27,9% (Tabela 3) da extensão predisposta no mapa litológico, enquanto unidades como a Formação Morro do Carneiro, Suíte Gameleira tiveram sua extensão aumentada em aproximadamente 170%. As

unidades referentes aos aluviões cenozoicos e diques cretáceos que correspondiam a estreitas ocorrências tiveram seu tamanho extrapolado durante a predição.

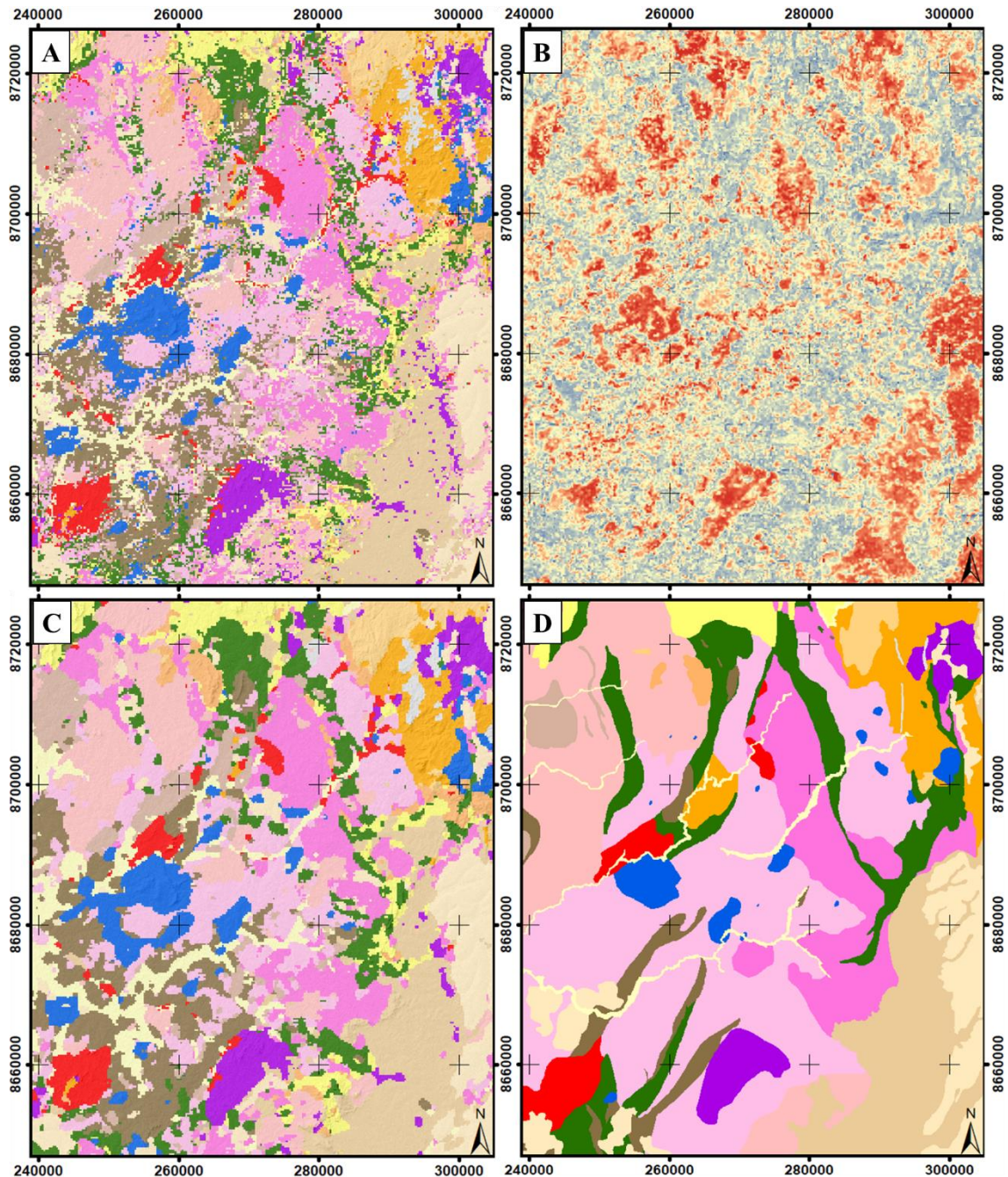


Fig. 9. Mapas resultantes do Teste A-09, classificação litológica realizada pelo algoritmo Random Forest com base nos parâmetros: TMI, ASA, MVI, K, eTh, eU, Kd, Alos e IM. Em A – Resultado bruto da classificação; B – Mapa probabilístico indicando em cores quentes e frias, altas e baixas probabilidades, respectivamente; C – Mapa preditivo litológico com uma filtragem por Majority; D – Mapa litológico de referência.

O teste A-09 teve seu resultado com base no uso de nove variáveis sendo elas mensuradas em importância pelo índice Gini, de maneira que a variável com maior importância o canal do potássio, seguido das seguintes variáveis: altimetria, índice

máfico, potássio anômalo, ASA, canal do Tório, slice do MVI, TMI e por último o canal do U.

Um dos produtos base do algoritmo *Random Forest* é a probabilidade de classificação de cada unidade, de maneira que a partir desta informação foi gerado um mapa probabilístico, ou mapa de incertezas (Fig. 9 – B). Este produto leva em consideração a maior probabilidade de ocorrência de uma determinada unidade para cada ponto, o mapa de incertezas nos traz informações de maneira visual para as regiões e unidades das quais o algoritmo apresentou uma maior assertividade. Para esses mapas optou-se por usar uma escala de cor de cores frias a cores quentes, baixas a altas probabilidades, respectivamente.

Tabela 3: Matriz de confusão comparando classes mapeadas e previsões do *Random Forest*. Os valores indicam a percentagem de amostras mapeadas presentes no mapa preditivo. Os destaques em verde, azul, amarelo, vermelho e cinza indicam um *recall** maior que 80%, 70%, 60%, 50% e inferior a 50%, respectivamente. O valor total no eixo X corresponde ao total de pontos que foi predito para a unidade, já o total no eixo Y o número de amostras de teste para a unidade.

MAP	Predicted																
	GrBam	DepA	Cdl	FCP	Dique	SGm	SGT	UnMonz	FMC	GrNat	SP	SPT	SQD	RIS	RAC	RAC-IB	SSB
GrBam	69.2	1.4	5.9	1.5	0.2	0.1	4.9	0.4	0.4	4.3	0.6	0.5	0.1	0.8	1.6	1.6	6.5
DepA	0.4	56.6	1.2	0.5	0.0	2.2	5.3	8.9	6.3	0.2	1.3	0.0	8.4	4.5	3.1	0.0	1.1
Cdl	2.5	1.9	67.2	4.4	0.1	4.3	3.5	0.5	9.2	0.2	0.1	0.0	0.2	3.7	1.9	0.0	0.3
FCP	6.8	1.9	5.3	27.2	0.4	5.2	7.5	1.0	9.5	9.1	1.9	0.4	3.8	9.8	2.9	5.2	1.9
Dique	0.0	0.0	0.0	0.0	76.0	0.0	0.0	0.0	0.0	0.0	16.0	4.0	0.0	0.0	0.0	0.0	4.0
SGm	0.0	1.1	2.9	1.3	0.8	77.7	1.6	0.8	3.1	2.1	1.7	0.1	0.7	5.1	0.3	0.6	0.3
SGT	5.9	3.8	3.4	10.7	0.2	1.2	46.9	1.7	1.3	4.2	0.7	0.2	4.0	6.5	5.8	2.6	0.7
UnMonz	2.1	9.9	0.9	2.8	0.0	0.6	2.3	60.5	4.2	1.0	1.8	0.2	2.3	3.5	1.3	1.8	4.6
FMC	0.7	10.6	2.0	4.6	0.2	9.1	0.3	4.8	49.4	1.7	0.1	0.0	4.9	10.0	1.2	0.3	0.2
GrNat	6.3	1.3	0.5	7.1	0.0	0.1	1.5	0.1	0.4	69.2	0.0	0.0	1.0	0.1	1.4	7.5	3.6
SP	1.9	1.7	1.1	1.5	8.9	5.4	2.8	0.6	1.6	0.5	41.0	14.2	4.1	4.0	0.8	5.8	4.1
SPT	0.7	0.0	0.0	0.1	1.2	0.1	0.6	0.0	0.0	0.0	8.5	85.5	0.0	0.0	0.7	1.9	0.8
SQD	0.2	5.3	0.2	3.3	0.0	0.1	6.4	0.7	7.6	0.5	0.0	0.0	59.8	6.4	9.4	0.0	0.0
RIS	3.4	8.1	4.2	8.6	0.6	6.0	6.9	2.1	13.6	1.9	1.5	0.0	7.9	27.9	6.1	0.7	0.6
RAC	2.2	8.6	1.8	2.7	0.1	0.8	11.2	1.1	5.4	3.9	0.1	0.4	7.2	7.7	42.1	3.3	1.3
RAC-IB	1.4	0.0	0.0	2.1	0.0	0.0	1.4	0.0	0.0	8.5	0.0	3.2	0.0	0.0	1.1	82.2	0.0
SSB	5.0	0.8	0.4	0.7	3.0	2.0	0.7	3.2	0.2	0.9	3.4	1.3	0.1	0.0	0.9	0.7	76.9
Σ	8224	5125	6470	6139	720	4035	8377	2504	6880	3561	2627	1561	4400	9821	7337	1971	3124

Buscando filtrar o mapa para ruídos de pequenas extensões, para uma visualização mais clara das unidades, aplicou-se o filtro *Majority* (Fig. 9 – C) para que o valor de cada pixel fosse determinado a partir do valor mais popular em uma janela de 8 por 8. A comparação com o mapa litológico original (Fig. 9 –D) mostra regiões em comum entre os dois mapas, mas também localidades nas quais a predição modificou a classificação inicial. Essa variação entre o mapa predito e o mapa original é visto na matriz de confusão (Tabela 3), onde é indicada a percentagem do mapa original que foi reconstruída pela predição, e as classes que foram classificadas de maneira distinta a original. A unidade

que teve a melhor reconstrução foi a Suíte Peraluminosa Tonalítica (SPT), com 85.5%. Para a análise mais minuciosa de como essas amostras foram classificadas calculou-se quanto do total de amostra originais foram classificadas de maneira distinta, e calculou-se a percentual para cada classe classificada em seu lugar (Fig. 10). A classe que mais teve em percentagem suas amostras classificadas como pertencentes a Suíte Gameleira foi a Formação Morro do Carneiro (FMC), já a Suíte Ribeirão Itaboca teve o maior número de amostras classificadas como Suíte Gameleira. O aumento de extensão considerável da FMC vem exatamente da reclassificação de amostras das unidades Fm. Córrego do Paiol, Suíte Ribeirão Itaboca (RIS), Suíte Ribeirão das Areias (RAC)

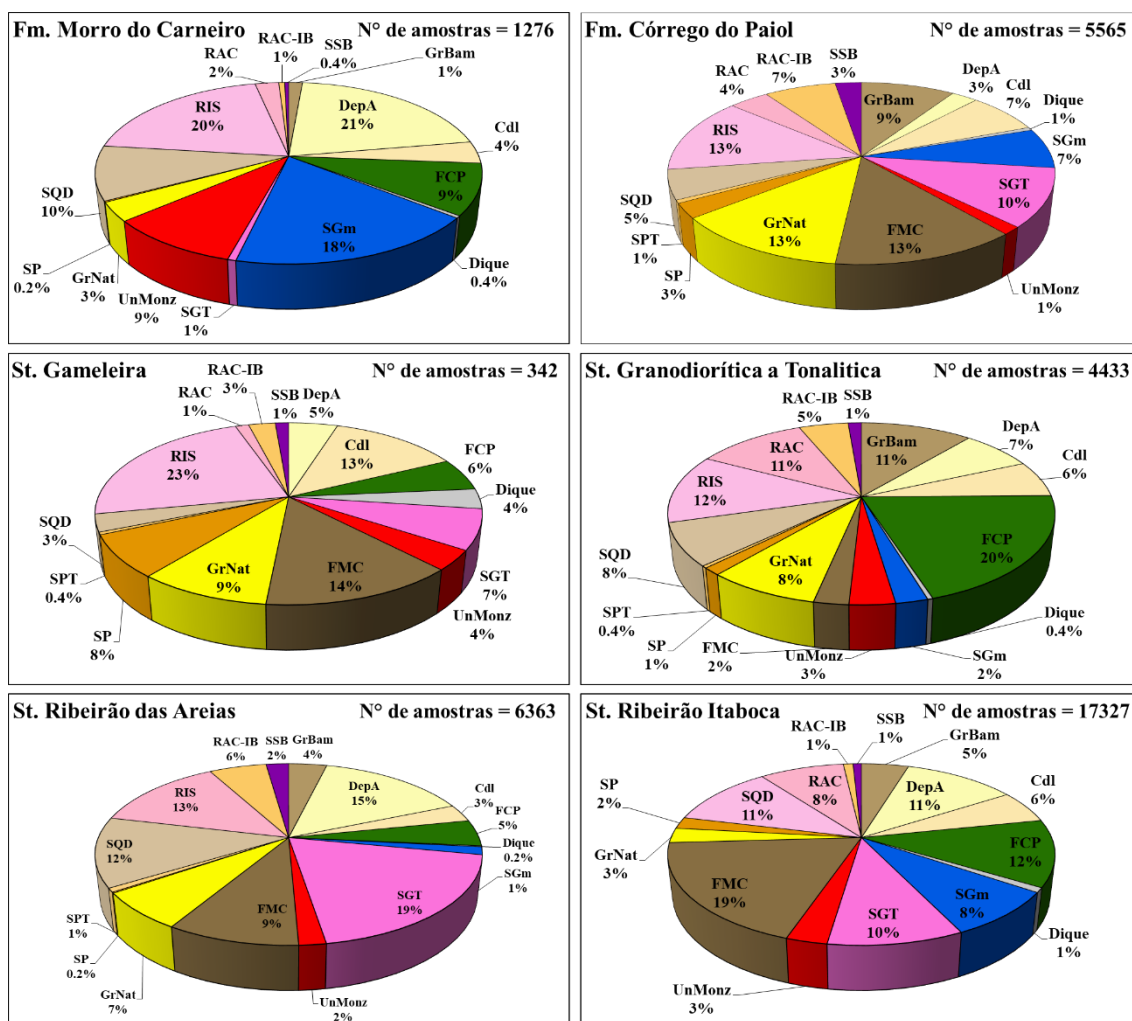


Fig. 10. Gráficos indicando o número total de amostras classificadas de maneira distinta da classe original, e o percentual de cada classe classificada no lugar.

Outra estratégia utilizada para a filtragem do mapa e o aumento da performance tanto dos modelos quanto da predição foi a utilização de uma das coordenadas UTM como um parâmetro de entrada. Dessa maneira para o mesmo conjunto de treino, utilizado nos testes apresentados acima, foi realizada a adição do *feature* referente a coordenada X-

UTM e gerado novos modelos e previsões (Fig. 11). Nota-se que o ruído teve uma redução significativa, bem como o aumento da performance do modelo e das previsões. No entanto, essa estratégia resultou em uma extrapolação da extensão de algumas unidades, como o caso da Suíte Granodiorítica Tonalítica (SGT), e a presença de alguns artefatos retilíneos.

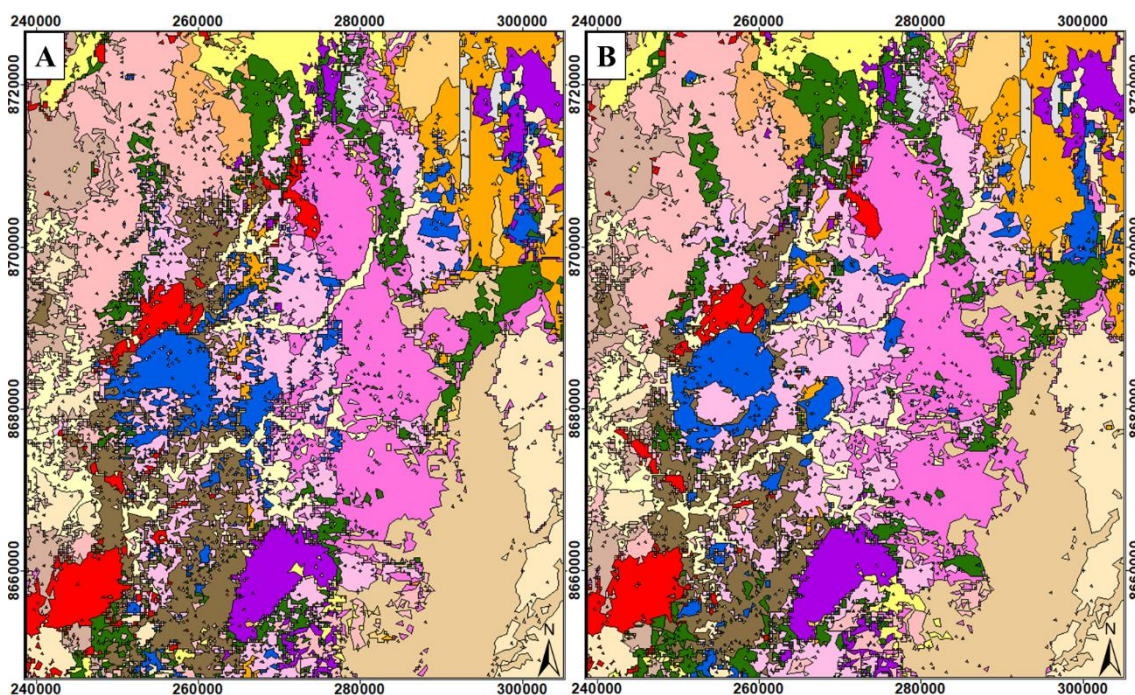


Fig. 11. Mapas resultantes da classificação litológica realizada pelo algoritmo Random Forest com base nos parâmetros: A – Teste A-02 (TMI, K, eTh, eU, Alos e X_UTM); B – Teste A-10 (TMI, ASA, MVI, K, eTh, eU, Kd, Alos, IM e X_UTM).

5.4 Lineamentos magnéticos e previsões geradas para a Suíte Gameleira

Os lineamentos magnéticos extraídos dos produtos aeromagnéticos referentes a derivada vertical, *Dz*, e a derivada *Tilt*, foram sobrepostas ao mapa probabilístico gerado via *Random Forest* somente para as probabilidades de ocorrência da Suíte Gameleira (Fig. 12 -A) e ao produto referente a derivada *Tilt* (Fig. 12 – B). O objetivo desse produto é avaliar se há relação entre as localidades de ocorrência dos corpos da suíte com determinadas estruturas, bem como sua direção preferencial. Na Fig. 12 é destacado que as regiões em que o algoritmo indicou mais alta probabilidade de ocorrência são “cortadas” ou estão próximos a lineamentos magnéticos, de direção predominantemente NE-SW e subordinadamente E-W e WNW-ESE.

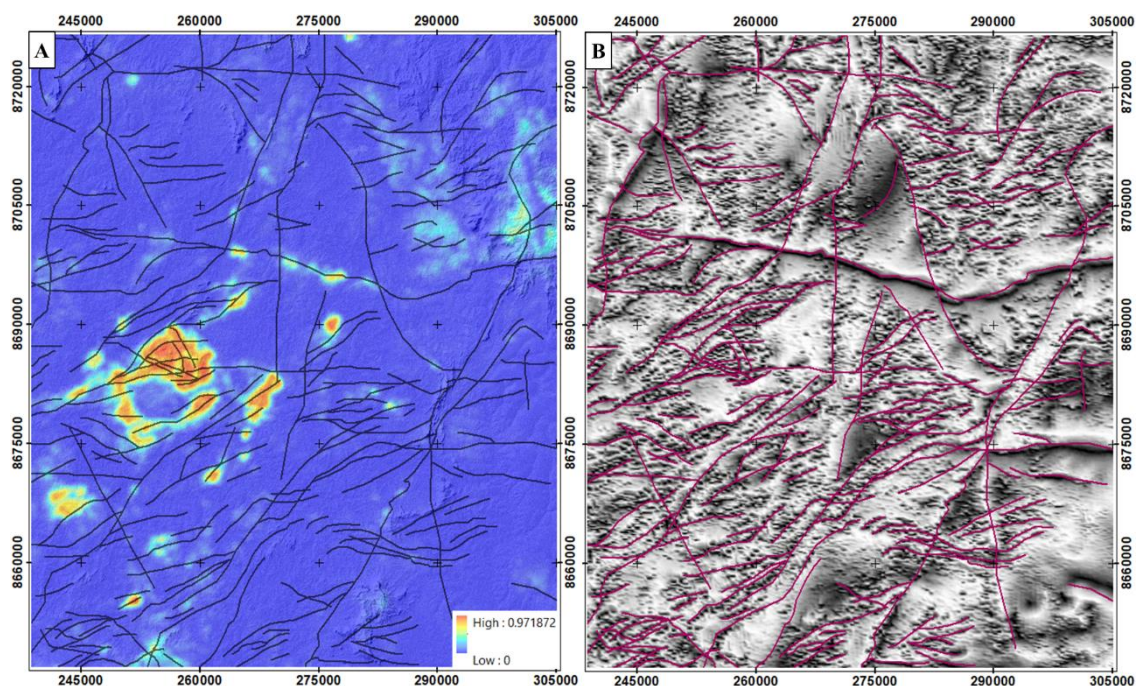


Fig. 12. Mapa indicando lineamentos magnéticos extraídos dos produtos Dz e TILT, em preto. Em escala de cor de azul a vermelho temos a probabilidade de ocorrência de corpos pertencentes a Suíte Gameleira.

5.5 Inversões do Vetor de magnetização e posicionamento de corpos máficos ultramáficos

De acordo com a literatura tem-se que as rochas que compõem a Suíte Gameleira possuem como característica alta assinatura magnética (Saboia & Meneghini, 2019), dessa maneira é esperado que as ocorrências de corpos dessa unidade venham atrelados a valores de intensidade magnética. Modelos de inversão utilizando o algoritmo *Magnetic Vector Inversion* (MVI) do *software Oasis Montaj*, foram gerados com diferentes tamanhos de células do mesh para cada área individualizada e tiveram o papel de espacializar as anomalias magnéticas existentes na área de estudo em profundidade, com finalidade de realizar verificação remota com base na propriedade física.

Na Fig. 13, foram geradas isosuperfícies com base nos valores de susceptibilidade magnética onde é possível relacionar os corpos indicados no mapa litológico preditivo com regiões em que há uma clara anomalia magnética característica situada logo abaixo dos limites mapeados, mas há também localidades em que não há uma anomalia magnética, esses locais são regiões onde o mapa probabilístico aponta uma probabilidade baixa de ocorrência da unidade. O inverso ocorre, há locais em que o mapa indica uma alta probabilidade de ocorrência e que não são mapeados como corpos pertencentes a

unidade, essas localidades são tidas como possíveis prospectos para a localização de novos corpos da unidade.

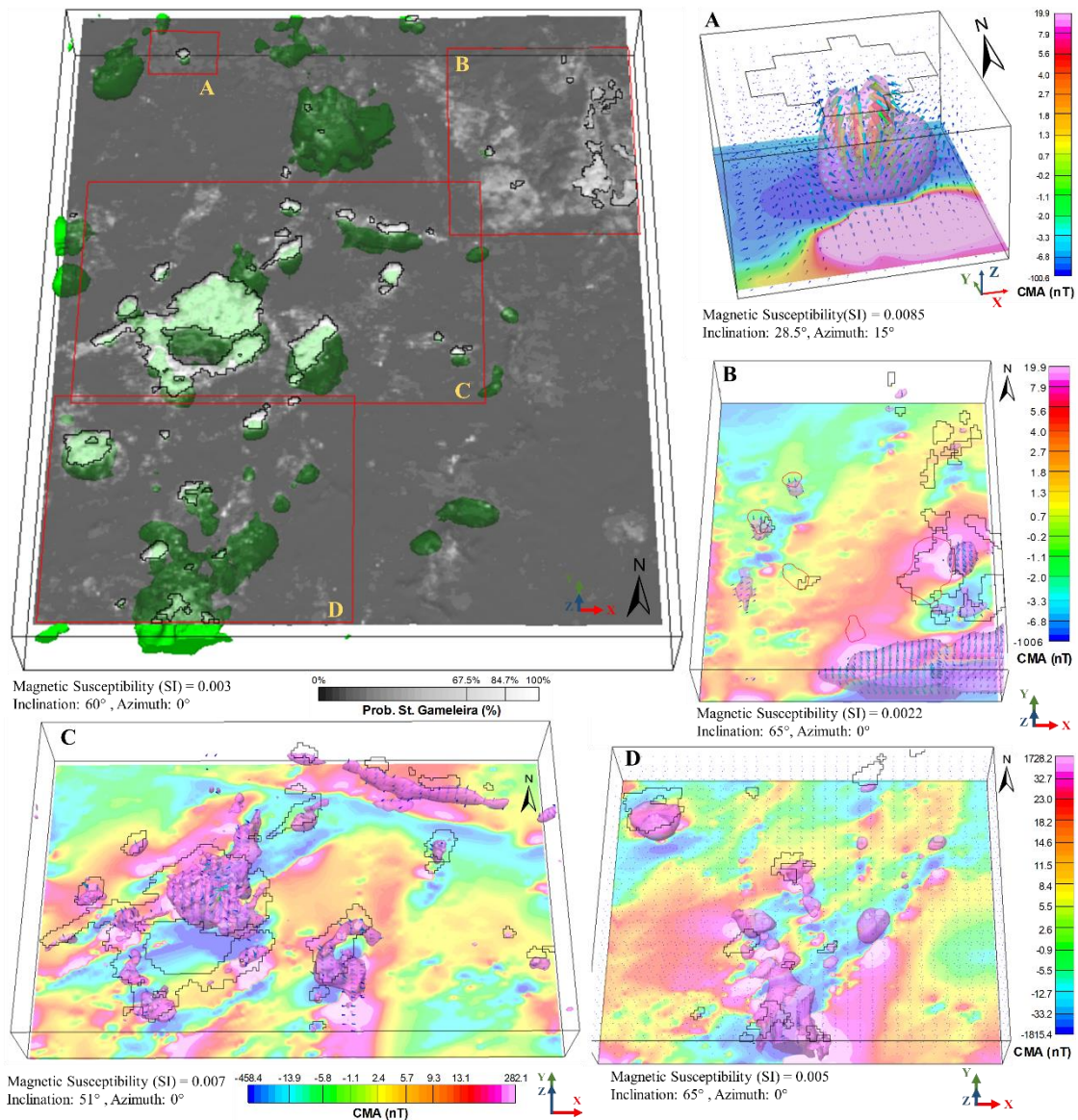


Fig. 13. Comparativo entre os produtos da inversão dos vetores de magnetização (MVI) e o produto do Random Forest referente a probabilidade de ocorrência da Suíte Gameleira, os polígonos em preto correspondem aos limites gerados pelo algoritmo. A – Produto referente a área de estudo como um todo, a variação de cinza a branco indica a menor para a maior probabilidade de ocorrência da unidade; B, C, D e E – Produtos referentes as subáreas destacadas em A, que apresentam maior probabilidade de ocorrência. Na base do modelo encontra-se o TMI.

6. Discussões

A aplicação do método de classificação *Random Forest* para a busca de novas ocorrências de corpos máficos ultramáficos pertencentes a Suíte Gameleira, se mostrou excelente ferramenta complementar para futuros trabalhos de campo. De maneira que os

mapas preditivos gerados, não correspondem em si a verdade geológica, mas podem indicar regiões passíveis de maior detalhamento.

6.1 Predições e classificação

Sabendo disso, este trabalho demonstrou que a partir de um conjunto de dados aerogeofísicos, a implementação do Random Forest foi capaz de classificar litologias com precisão de 55,8% com base em um mapa litológico prévio, utilizando somente 1,96% da base de dados como amostras de treino para o modelo. Os resultados obtidos quando comparados a trabalhos que usaram técnicas semelhantes não estão tão distantes, como os trabalhos de Cracknell *et al.*, (2014a), Harris & Grunsky (2015) e Kuhn *et al.*, (2018) que utilizaram metodologias semelhantes e obtiveram acurácia de 78%, 57.8% e 76%, respectivamente. Já Yu *et al.*, (2012) obteve classificação de 62.2%, utilizando SVM. Por mais que a precisão alcançada não seja tão alta, os resultados obtidos são melhores que nenhuma informação geológica, de maneira que é importante ressaltar os diferentes conjuntos de dados utilizados em cada estudo, e a complexidade de cada área. A área de estudo apresenta uma complexidade radiométrica e magnética, em que os padrões vistos nos produtos muitas vezes não correspondem aos contatos indicados no mapa litológico. Algumas dessas áreas apresentam alta incerteza de classificação, um dos motivos prováveis é a proximidade dos valores de propriedade física entre as unidades, uma vez que a variação de teores de silício entre rochas ácidas à ultrabásicas, tem como resultante o aumento na concentração de radioelementos (Dickson & Scott, 1997; Wilford & Minty, 2006), e na área de estudo temos um terreno granito-gnáissico com assembleias minerais similares. De forma semelhante a quantidade de minerais (ricos em Fe e Mg), tendem a possuir um maior valor de susceptibilidade magnética (Dentith & Mudge, 2014; Schon, 2011), dessa maneira na área de estudo há ocorrência de rochas máficas atreladas as unidades do Grupo Riachão do Ouro (rochas vulcânicas máficas e formações feríferas bandadas) e a Suíte Gameleira. O trabalho de campo evidenciou um grupo diverso de rochas máficas e ultramáficas pertencentes a Suíte Gameleira, que indicam uma variação de concentração e tipos de minerais máficos de maneira que até mesmo dentro da unidade podemos ter diferentes tipos de assinaturas.

Na Fig. 8 podemos notar os diferentes mapas preditivos gerados a partir da variação do dado magnético de entrada, como a Suíte Gameleira apresenta a maior mediana no que diz respeito a intensidade magnética (Fig. 6 – D) o intuito destes experimentos foi o de entender qual seria a melhor entrada magnética para a localização

da unidade com maior precisão. A composição dos *features* de entradas magnéticas TMI, ASA e o MVI foi o melhor resultado de maneira que o resultado do teste com o ASA e o MVI apresentaram resultados bem semelhantes em comparação com o utilizando o TMI que individualizou mais corpos, teve uma geometria do corpo maior central aparentemente estirada e os corpos a nordeste tiveram uma melhor delineação, uma vez que nos demais testes eles apresentaram um formato alongado para norte, justamente por haver uma fonte magnética com sentido norte-sul, bem discretizada nos produtos ASA e *slice* do MVI (Fig. 3 – B e C).

No teste A-09 pode-se ver a partir da matriz de confusão (Tabela 3) que as unidades com melhor reconstrução, quando comparado com o mapa litológico foram: Suíte Peraluminosa-Tonalítica (SPT) com 85.5%, os corpos intrusivos Suíte Ribeirão das Areias (RAC-IB) com 82.2%, Suíte Gameleira (SGm) com 77.7% e a Suíte Serra do Boqueirão (SSB) com 76.9%. Esses resultados são possivelmente relacionados a sua pequena distribuição espacial e assinatura de propriedade física restrita da unidade como visto na Fig. 7, com exceção da Suíte Gameleira que deve ser ponderado o seu contraste de propriedade física.

As unidades que tiveram sua extensão consideravelmente modificada, relativamente distantes do mapa litológico inicial, destacam-se as unidades Formação Córrego do Paiol (FCP) com 27.2%, Ribeirão de Itaboca (RIS) com 27.9%, Suíte Peraluminosa (SP) com 41%, Suíte Ribeirão das Areias (RAC) com 42.1% e a Suíte Granodiorítica Tonalítica (SGT) com 46.9%. A baixa precisão na classificação dessas unidades pode ser relacionada a motivos como a grande variância na assinatura das amostras, ampla cobertura espacial e possivelmente classificação equivocada no mapa litológico original. Nota-se que três unidades no mapa original (RIS, RAC e SGT) correspondem a aproximadamente 53% da extensão total da área, possuindo muito mais pontos amostrais e quando analisada sua assinatura radiométrica e magnética é visto um conjunto de assinaturas difusas, com presença de inúmeros *outliers*, o que pode ser um dos motivos da baixa classificação geral da predição.

Quando analisado por unidade, a frequência com que uma classificação diferente da esperada foi realizada (Fig. 10), algumas unidades foram classificadas com grande frequência, como é o caso da Formação Morro do Carneiro (FMC) que é composta por uma ampla variedade de litotipos (metassedimentos terrígenos e químicos, metavulcânicas félsica, intermediárias e máficas) de forma a ter uma assinatura tanto

radiométrica como magnética com ampla variância. A FMC foi a mais frequentemente classificada de maneira distinta da classe original para as unidades FCP, SGM e Ribeirão de Itaboca, dessa maneira tendo sua extensão espacial significativamente ampliada.

Quando adicionado o *feature* referente a coordenada espacial como sugerido por Cracknell, 2014a e Costa, 2019, houve melhora significativa na performance do modelo e da predição, no entanto, quando espacializada as informações algumas feições de cunho não geológico e uma extrapolação no tamanho de alguns limites de unidades, essa relação traz o questionamento se por vezes o aumento da performance do algoritmo vem atrelada a preservação e manutenção do sentido físico do dado.

6.2 Verificação indireta ou remota

Quando individualizada a probabilidade de ocorrência para a Suíte Gameleira através do mapa de incertezas, nota-se que as ocorrências de corpos com maior probabilidade coincidem com os lineamentos magnéticos extraídos dos produtos magnéticos DZ e TILT. Esta informação é uma das utilizadas para a averiguação de coerência do resultado, tendo em vista que até então essa unidade geológica possui origem incerta e não há trabalhos que infiram um condicionamento tecto-estutural para a Suíte Gameleira. Dois *trends* de lineamentos perpassam esses corpos podendo ser ou de direção NE ou com direção, aproximadamente, E-W.

Uma outra abordagem para a inferência com maior certeza destes corpos foi o uso dos modelos de inversão dos vetores de magnetização, uma vez que esses corpos possuem uma assinatura em quase todas as suas ocorrências de remanência magnética, marcada pela inversão dos polos no TMI. Essa característica para a região de estudo tem um caráter importante, mas a avaliação deve ser feita de forma cuidadosa uma vez que há na região outras unidades que podem apresentar remanência magnética como os BIF's da unidade Formação Morro do Carneiro e as rochas vulcânicas máficas-ultramáficas da Formação Córrego do Paiol, ambas formações fazem parte do Grupo Riachão do Ouro com idades arqueanas-paleoproterozoicas podendo terem passados por eventos de remagnetização e alteração do sentido do campo magnético local.

O corpo ilustrado situado a noroeste da área de estudo indicado pelo algoritmo, foi um dos averiguados pela inversão e pela análise de lineamentos e anomalia magnética, o mesmo é individualizado na Fig. 13 - A, podemos notar perturbação e variação de intensidade e sentido dos vetores de magnetização no corpo, e anomalia de polos

invertidos clássica do hemisfério sul. Este foi um dos corpos visitados durante a etapa de campo e apresentou rochas como hornblenda gabros e leucogabros, sendo uma das predições realizadas pelo algoritmo que não estava inserido dentro do mapa geológico inicial.

7. Conclusões

A aplicação do *Random Forest* foi eficiente como classificador de litologias, gerando produtos indicativos de possíveis localidades para ocorrências de corpos máfico-ultramáficos da Suíte Gameleira. Além disso, indicou regiões que são passíveis de um melhor detalhamento cartográfico, com o aumento da escala de mapeamento, como, por exemplo, a ampla extensão pertencente ao embasamento granito-gnaissico.

No uso do classificador supervisionado *Random Forest* a escolha dos parâmetros de entrada são essenciais e vão refletir o resultado da predição. O presente trabalho elucida como os modelos gerados para diferentes tipos de entrada de dados magnéticos são distintos. Para a área de estudo, a utilização da combinação de *features* magnéticos resultou na melhor aderência ao modelo, possivelmente pela complexidade magnética da área (presença de anomalias em que há predominância da magnetização induzida frente o vetor de magnetização remanente e vice-versa). O uso dos *features* de coordenadas resultou em um aumento na performance do modelo e de suas predições, no entanto generalizaram e extrapolaram os limites em algumas localidades da área, chegando a apresentar artefatos retilíneos.

O uso de técnicas de verificação foi essencial para corroborar com a predição do modelo, bem como aumentar a confiabilidade de suas predições, em específico para os corpos pertencentes a Suíte Gameleira. Os dados de campo confirmaram a ocorrência de rochas pertencentes a Suíte Gameleira, na porção noroeste da área, indicada na predição, e confirmam uma variação litológica da unidade, que pode corresponder a variância e presença de *outliers* indicados nos *gráficos BoxPlots* (Fig. 7). A investigação de um controle tectônico para a alocação dos corpos da Suíte Gameleira através do uso de lineamentos magnéticos extraídos da primeira derivada e da derivada Tilt, indicaram a presença de lineamentos com direção NE-SW que perpassam ou estão próximos dos corpos que o algoritmo indicou uma maior probabilidade de ocorrência, com alguns já confirmados via dados de campo. Por fim, a inversão dos vetores de magnetização (MVI), técnica proposta para verificação remota de corpos máfico-ultramáficos, foi essencial,

uma vez que confirmaram anomalias magnéticas em profundidade, com geometrias circulares, que validam o modelo de corpos acamadados. As variações da susceptibilidade magnética dos produtos de inversão marcam maior intensidade em corpos onde há condicionante tectônica, e em alguns casos a baixa intensidade é relacionada a presença de rochas de máficas como uma mineralogia magnética menos expressiva, leucogabros, mapeados na subárea B (Fig. 13).

Mediante a isso, o uso de técnicas de aprendizado de máquina, *Random Forest*, para a predição litológica mostrou-se uma ferramenta complementar útil para a prospecção de novos corpos máfico-ultramáficos.

Capítulo 6 – Considerações Finais

Por meio do algoritmo Random Forest, esta dissertação teve como objetivo localizar novas ocorrências de corpos máfico-ultramáficos pertencentes a Suíte Gameleira. Os produtos finais do algoritmo indicaram novas regiões nas quais há uma boa probabilidade de novos corpos dessa Suíte. O algoritmo teve como base de dados, múltiplos produtos gerados a partir dos dados aerogeofísicos (magnetométricos e radiométricos), dados altimétricos (satélite Alos sensor Palsar) e geológicos (folhas geológicas e mapas de referência encontrados em trabalhos prévios). As predições finais geraram uma classificação de acurácia com média de aproximadamente 50%, no entanto com recuperação maior que 77% para a classe referente a Suíte Gameleira.

Concluimos que a qualidade dos dados de entrada, principalmente o mapa litológico de referência, é fundamental, uma vez que a entrada do controle litológico da área foi responsável por classificar todas as amostras da base de dados quanto a que classe geológica cada uma pertencia. Visto que o mapa de entrada está na escala 1:250000, uma das interpretações para as diferenças do mapa original frente ao predito é o de um aumento de detalhe em determinadas regiões. Isso é observado nos gráficos *BoxPlot*, onde há uma presença significativa de outliers estes sendo interpretados como: (i) classificação errônea devido a maior escala do mapa de entrada; (ii) variações faciológicas dentro da unidade, não há uma uniformidade na assinatura radiométrica e magnética em algumas unidades; (iii) ocorrências anômalas, hidrotermalismo ou mineralizações.

Propomos ainda, o uso de ferramentas de verificação remota para a localização de corpos máfico-ultramáficos, sendo realizada de três maneiras neste trabalho: (i) verificação da polaridade da anomalia magnética; (ii) correlação com estruturas de lineamentos magnéticos; (iii) avaliação do corportamento em profundidade da anomalia utilizando métodos de inversão magnética. Essas técnicas de verificação remota aumentam a credibilidade do modelo predito, corroborando com alvos ou não. A verificação direta ainda é essencial, no entanto pode ser de difícil realização devido a custos e áreas de difícil acesso (propriedades privadas, locais remotos de difícil acesso).

6.1 Sugestões para trabalhos futuros

- ❖ Mapeamento de detalhe na área de estudo, sugere-se 1:100000. Com a cartografia de detalhe realizada, uma nova etapa de rotulo das amostras seria realizada e o resultado comparado;
- ❖ Utilização de produtos decorrentes do processamento de imagens de satélite, raízes de bandas, índices de vegetação, entre outros produtos, como features de entrada buscando aumentar o conjunto de informações de cada amostra;
- ❖ Análise estatística minuciosa, comparando o agrupamento de propriedades físicas inicial com o predito:
 - Entender o comportamento dos *outliers* pós classificação.
- ❖ Datação e realização de química mineral das rochas da Suíte Gameleira, para caracteriza-las e situa-las quando a seu ambiente tectônico de formação e geocronologia;
- ❖ Caracterização da assinatura de propriedades físicas frente a variações faciológicas da Suíte Gameleira, procurando individualizar assinaturas de rochas máficas de rochas ultramáficas;
- ❖ Quanto ao caráter econômico da unidade sugere-se a realização de furos de sondagem rotativa, para que seja averiguada a presença ou não de mineralizações de cobre ou níquel que seja suficiente para a implementação de uma frente de lavra.

Referências Bibliográficas

- Abdallah, S. (2015). Geologia e Geoquímica do Grupo Riachão do Ouro na Folha Arraias: Evidências de Arco Magmático Paleoproterozoico. *Geochimica Brasiliensis*, 29(2), 100-100.
- Abdallah, S., Meneghini, P. F. V. B., (2017). Geologia e recursos minerais da folha Arraias, SD. 23-VA, estado do Tocantins. CPRM (in português).
- Adams, J. A., & Gasparini, P. (2013). *Gamma-ray spectrometry of rocks*. Elsevier.
- Aisengart, T., 2013. 3D Inversion of Magnetic Data at Low Magnetic Latitudes. Thirteenth International Congress of the Brazilian Geophysical Society, Rio de Janeiro.
- Aisengart, T., 2015. Qualitative and quantitative magnetization vector inversion applied to the Pirapora anomaly. In 14th International Congress of the Brazilian Geophysical Society & EXPOGEF, Rio de Janeiro, Brazil, pp. 474-479, Brazilian Geophysical Society.
- Aitkenhead, M. J., Coull, M. C., & Dawson, L. A. (2014). Predicting Sample Source Location from Soil Analysis Using Neural Networks. *Environmental Forensics*, 15(3). <https://doi.org/10.1080/15275922.2014.930764>
- Alkmim, F. F., & Martins-Neto, M. A. (2012). Proterozoic first-order sedimentary sequences of the São Francisco craton, eastern Brazil. *Marine and Petroleum Geology*, 33(1), 127-139.
- Almeida, F. F. M., Hasui, Y., Brito Neves, B. B. & Fuck, R. A. 1981. Brazilian structural provinces: an introduction. *Earth-Science Reviews*, 17, 1–29.
- de Alvarenga, C. J. S., Dardenne, M. A., Vieira, L. C., Martinho, C. T., Guimarães, E. M., Santos, R. V., & Santana, R. O., 2012. Estratigrafia da borda ocidental da Bacia do São Francisco.
- Angelov, P. P., Soares, E. A., Jiang, R., Arnold, N. I., & Atkinson, P. M. (2021). Explainable artificial intelligence: an analytical review. *Wiley Interdisciplinary Reviews: Data Mining and Knowledge Discovery*, 11(5), e1424.
- Araújo-Filho, J.O. 2000. The Pireneus Syntaxis: an example of the intersection of two Brasiliano fold-thrust belts in central Brazil and its implications for the tectonic evolution of western Gondwana. *Rev. Bras. Geocienc.* 30:144-148.
- Austin, J., Hillan, D., & Foss, C. (2019). Remanent magnetization mapping: A tool for greenfields magmatic Ni-Cu-PGE exploration undercover: Part 2. *Ore Geology Reviews*, 109, 290-302.
- Baranov, V. (1957). A new method for interpretation of aeromagnetic maps: pseudo-gravimetric anomalies. *Geophysics*, 22(2), 359-382.
- Barbosa, R.D., & Pereira, J.G., 2013. Inversão 3D de dados magnéticos na região de Mara Rosa-Goiás, Brasil, utilizando Geosoft VOXI. In 13th International Congress of the Brazilian Geophysical Society & EXPOGEF, Rio de Janeiro, Brazil, pp. 520-525, Society of Exploration Geophysicists and Brazilian Geophysical Society.

- Barnes, S. J., Malitch, K. N., & Yudovskaya, M. A. (2020). Introduction to a special issue on the Norilsk-Talnakh Ni-Cu-platinum group element deposits. *Economic Geology*, 115(6), 1157-1172.
- Bhattacharyya, B. K. (1965). Two-dimensional harmonic analysis as a tool for magnetic interpretation. *Geophysics*, 30(5), 829-857.
- Bizzi, L. A., Schobbenhaus, C., Vidotti, R. M., & Gonçalves, J. H. (2003). *Geologia, Tectônica e Recursos Minerais do Brasil: texto, mapas e SIG*. CPRM.
- Blakely, R. J. (1996). *Potential theory in gravity and magnetic applications*. Cambridge university press.
- Blum, M.L.B., 1999. Processamento e interpretação de dados de geofísica aérea no Brasil Central e sua aplicação à Geologia Regional e à Prospecção Mineral. Tese de Doutorado. Universidade de Brasília. 229 p.
- Borges, M. S., Costa, J. B. S., Hasui, Y., 1998. Lito-estratigrafia da sequência metavulcanosedimentar de Dianópolis-Almas, sudeste do Tocantins. *Geociências*17(1), 61-79 (in portuguese).
- Borges, M. S., Costa, J. B. S., Hasui, Y., 1999. A Estruturação da Sequência Metavulcanosedimentar de Almas-Dianópolis, Sudeste de Tocantins. *Anais da Academia Brasileira de Ciências* 71(4 PART 1), 697-716 (in portuguese).
- Breiman L., Friedman J., Olshen R.A., Stone C.J. 1984. Classification and regression trees, the wadsworth statistics and probability series. Wadsworth International Group, Belmont California, 358 p
- Breiman, L (1996) Bagging predictors, *Machine Learning* 24:123-140.
- Breiman, L. (2001). Random forests. *Machine learning*, 45, 5-32.
- Brito Neves, B.B., Cordani, U.G. (1991). Tectonic evolution of South America during the Late Proterozoic. In: STERN, R.J., Van Schmus, W.R. Crustal evolution in the late Proterozoic. *Precambrian Research*, 53, 23-40.
- Brito Neves, Benjamim Bley (2002). Main stages of the development of the sedimentary basins of South America and their relationship with the tectonics of supercontinents. *Gondwana Research*, v. 5, n. 1, p. 175-196.
- Burkov, A. (2019). *The hundred-page machine learning book* (Vol. 1, p. 32). Quebec City, QC, Canada: Andriy Burkov.
- Castro, D. L., Fuck, R. A., Phillips, J. D., Vidotti, R. M., Bezerra, F. H., Dantas, E. L., 2013. Crustal structure beneath the Paleozoic Parnaíba Basin revealed by airborne gravity and magnetic data, Brazil. *Tectonophysics*, 614, 128-145.
- Cawthorn, R. G. (2010). The platinum group element deposits of the Bushveld Complex in South Africa. *Platinum Metals Review*, 54(4), 205-215.
- Chai, G., & Naldrett, A. J. (1992). Characteristics of Ni-Cu-PGE mineralization and genesis of the Jinchuan deposit, northwest China. *Economic Geology*, 87(6), 1475-1495.
- Chapman, S., & Bartels, J. (1940). *Geomagnetism*, vol. II: Analysis of the data, and physical theories. *Geomagnetism*.

- Cordeiro, P. F. O., Oliveira, C. G., Della Giustina, M. E. S., Dantas, E. L., Santos, R. V., 2014. The Paleoproterozoic Campinorte Arc: Tectonic Evolution of a Central Brazil pre-Columbia Orogeny. *Precambrian Research*, 251, 49-61.
- Correia Filho, F.C.L. Sá, A.M., (1980). Projeto Natividade. Goiânia, DNPM/CPRM. Relatório técnico, p. 120 (in portuguese).
- Costa M.T. da & Branco (1961). Roteiro para a excursão Belo Horizonte - Brasília. In: SBG, Congresso Brasileiro de Geologia, 14, Belo Horizonte, EEUFMG-Instituto de Pesquisas Radioativas, Publicação 15, 25 p.
- Costa, I. S. L., Serafim, I. C. C. D. O., Tavares, F. M., & Polo, H. J. D. O. (2020). Uranium anomalies detection through Random Forest regression. *Exploration Geophysics*, 51(5), 555-569.
- Costa, I. S. L., Tavares, F. M., & de Oliveira, J. K. M. (2019). Predictive lithological mapping through machine learning methods: a case study in the Cinzento Lineament, Carajás Province, Brazil. *Journal of the Geological Survey of Brazil*, 2(1), 26-36.
- Costa, J. B.S., (1985). Aspectos Lito-Estruturais e Evolução Crustal da Região Centro-Norte de Goiás. Unpublished PhD. Thesis. Belém, UFPA. Centro de Geociências, p. 209 (in portuguese).
- Costa, L.A.M., Portella, A.C.P. and others, 1976. Projeto Leste do Tocantins-Oeste do Silo Francisco (unpubl.). Relatório Final, Fase V. Prospec -- D.N P.M---C.P R.M_, 300 pp.
- CPRM, (2000). Geologia e resultados prospectivos da área Barra do Gameleira, Tocantins.
- CPRM, (2006). Projeto Aerogeofísico Tocantins: Relatório final do levantamento e processamento dos dados magnetométricos e gamaespectrométricos.
- Cracknell M., Reading, A. (2014a). Geological mapping using remote sensing data: a comparison of five machine learning algorithms, their response to variations in the spatial distribution of training data and the use of explicit spatial information. *Computers & Geosciences*, 63, 22-33. <https://doi.org/10.1016/j.cageo.2013.10.008>
- Cracknell, M. J., & Reading, A. M. (2013). The upside of uncertainty: Identification of lithology contact zones from airborne geophysics and satellite data using random forests and support vector machines. *Geophysics*, 78(3), WB113-WB126.
- Cracknell, M. J., Reading, A. M., & McNeill, A. W. (2014b). Mapping geology and volcanic-hosted massive sulfide alteration in the Hellyer–Mt Charter region, Tasmania, using Random Forests™ and Self-Organising Maps. *Australian Journal of Earth Sciences*, 61(2), 287-304.
- Cruz, E. L. C. C; Kuyumjian, R. M; Boa Ventura, G. R. 2003. low-k calc-alkaline granitic series of southeastern tocanins state: chemical evidence for two sources for the granite gneissic complexes in the paleoproterozoic Almas-Dianópolis. *Revista Brasileira de Geociências* 33(2):125-136.
- Cruz, E.L.C.C.; Kuyumjian R.M. 1999. Mineralizações auríferas filoneanas do terreno granitogreenstone do Tocantins. *Rev. Bras. Geoc.*, 29(3), 291-298 (in portuguese).

Cruz, E.L.C.C.; Kuyumjian R.M., 1998. The geology and tectonic evolution of the Tocantins granite-greenstone terrane, Almas-Dianópolis region, Tocantins State, Central Brasil. *Revista Brasileira de Geociências* 28(2), 173-182.

Danni, J. C., & Teixeira, N. A. (1981). Características e sistematização das associações de rochas máficas e ultramáficas pré-cambrianas do estado de Goiás. *Simp. Geologia. Centro-Oeste*, 1, 376-401.

Dardenne, M.A. Síntese sobre a estratigrafia do Grupo Bambuí no Brasil Central. In: 30º Congresso Brasileiro de Geologia, 1978, Recife. *Anais*, v. 2. p. 597–610, (1978).

Dardenne, M.A., Giustina, M.E.S.D., Saboia, A.M., Bogossian, J., (2009). Datação Geocronológica U–Pb da Sequência Vulcânica de Almas, Tocantins. In: SBG, *Simp.Geol. Centro-Oeste*, 11, *Anais*, 1 CD-Rom (in portuguese).

Dardenne, Marcel Auguste; Saboia, André Menezes. Litoestratigrafia do grupo natividade na região de Natividade-Pindorama, sudeste do estado do Tocantins. In: ROSA-COSTA, Lúcia Travassos; KLEIN, Evandro Luiz; VIGLIO, Eduardo Paim (Org.). *Contribuições à geologia da Amazônia*. Belém: SBGeo-NO, (2007). p. 29-38. ISBN 9788588692046.

Dentith, M., & Mudge, S. T. (2014). *Geophysics for the mineral exploration geoscientist*. Cambridge University Press.

Dickson, B. L., & Scott, K. M. (1997). Interpretation of aerial gamma-ray surveys-adding the geochemical factors. *AGSO Journal of Australian Geology and Geophysics*, 17(2), 187-200.

e Silva, J. M., Ferreira Filho, C. F., Bühn, B., & Dantas, E. L. (2011). Geology, petrology and geochemistry of the “Americano do Brasil” layered intrusion, central Brazil, and its Ni–Cu sulfide deposits. *Mineralium Deposita*, 46, 57-90.

Ellis, R., de Wet, B., & Macleod, I. M. 2012. Inversion of magnetic data from remanent and induced sources: Presented at the 22nd International Geophysical Conference. 10.1071/ASEG2012ab117

Ferreira Filho, C. F., Pimentel, M. M., de Araujo, S. M., & Laux, J. H. (2010). Layered intrusions and volcanic sequences in Central Brazil: geological and geochronological constraints for Mesoproterozoic (1.25 Ga) and Neoproterozoic (0.79 Ga) igneous associations. *Precambrian Research*, 183(3), 617-634.

Ferreira, F. D. F., Souza, J., Bongioio, A. DE B. E S., Castro, L. G. (2013). Enhancement of the total horizontal gradient of magnetic anomalies using the tilt angle. *Geophysics*, 78: 33- 41.

Fuck, R. A. (1994). A Faixa Brasília e a compartimentação tectônica na Província Tocantins. SBG, *Simpósio de Geologia do Centro-Oeste*, 4, 184-187.

Fuck, R.A. Pimentel, M.M. Soares J.E., Dantas, E. L. (2005). Compartimentação da Faixa Brasília. In: *Simpósio de Geologia do Centro-Oeste*, 9, Goiânia. *Anais*. p.26-27.

Fuck, R.A., Dantas, E.L., Pimentel, M.M., Botelho, N.F., Armstrong, R., Laux, J.H., Junges, S.L., Soares, J.E., Praxedes, I.F., (2014). Paleoproterozoic crust-formation and reworking events in the Tocantins Province, central Brazil: A contribution for Atlantica supercontinent reconstruction. *Precambrian Research* 244, 53-74.

- Fuck, R.A., Pimentel, M.M., Alvarenga, C.J., Dantas, E.L., 2017. The northern Brasília belt. In São Francisco Craton, Eastern Brazil , 205-220. Springer Cham.
- Galar, M., Fernandez, A., Barrenechea, E., Bustince, H., & Herrera, F. (2011). A review on ensembles for the class imbalance problem: bagging-, boosting-, and hybrid-based approaches. *IEEE Transactions on Systems, Man, and Cybernetics, Part C (Applications and Reviews)*, 42(4), 463-484.
- Giovanardi, T., Girardi, V. A., Correia, C. T., Tassinari, C. C., Sato, K., Cipriani, A., & Mazzucchelli, M. (2017). New U-Pb SHRIMP-II zircon intrusion ages of the Cana Brava and Barro Alto layered complexes, central Brazil: constraints on the genesis and evolution of the Tonian Goiás Stratiform Complex. *Lithos*, 282, 339-357.
- Gnojek, I., & Přichystal, A. (1985). A new zinc mineralization detected by airborne gamma-ray spectrometry in northern Moravia (Czechoslovakia). *Geoexploration*, 23(4), 491-502.
- Gorayeb, P. D. S., Costa, J. B. S., LEMos, R. L., Gama Jr, T., Bemerguy, R. L., & Hasui, Y. O. C. I. T. E. R. U. (1988). O pré-cambriano da região de Natividade, GO. *Revista Brasileira de Geociências*, 18(4), 391-397.
- Grauch, V. J. S., & Cordell, L. (1987). Limitations of determining density or magnetic boundaries from the horizontal gradient of gravity or pseudogravity data. *Geophysics*, 52(1), 118-121.
- Harris, J. R., & Grunsky, E. C. (2015). Predictive lithological mapping of Canada's North using Random Forest classification applied to geophysical and geochemical data. *Computers & geosciences*, 80, 9-25.
- Harris, J. R., Schetselaar, E., & Behnia, P. (2012). *Remote predictive mapping: an approach for the geological mapping of Canada's arctic*. INTECH Open Access Publisher.
- Hasui, Yociteru; Costa, João Batista Sena; Saad, Antônio Roberto; Campanha, Vilma Alves. O Grupo Natividade e sua correlação com o Grupo Bambuí. *Geociências*, São Paulo, p. 299-316, 1990. (Edição especial).
- Hill, E. J., Fabris, A., Uvarova, Y., & Tiddy, C. (2021). Improving geological logging of drill holes using geochemical data and data analytics for mineral exploration in the Gawler Ranges, South Australia. *Australian Journal of Earth Sciences*, 1-27.
- IAEA - International Atomic Energy Agency. 2003. Guidelines for radioelement mapping using gamma ray spectrometry data. IAEA TECDOC, 1363, 184 p.
- Johnson, A., & Aisengart, T., 2014. Interpretation of magnetic data at low magnetic latitudes using magnetization vector inversion. *Journal of Geophysics*, 35(3), 91-96.
- Jung, D., Choi, Y. (2021). Systematic Review of Machine Learning Applications in Mining: Exploration, Exploitation, and Reclamation. *Minerals* 2021, Vol. 11, Page 148, v. 11, n. 2, p. 148.
- Kotsiantis, S. B., Zaharakis, I. D., & Pintelas, P. E. (2006). Machine learning: a review of classification and combining techniques. *Artificial Intelligence Review*, 26, 159-190.

Kuhn, S., Cracknell, M. J., & Reading, A. M. (2018). Lithologic mapping using Random Forests applied to geophysical and remote-sensing data: A demonstration study from the Eastern Goldfields of Australia. *Geophysics*, 83(4), B183-B193.

Kuhn, S., Cracknell, M. J., & Reading, A. M. (2019). Lithological mapping in the Central African Copper Belt using Random Forests and clustering: Strategies for optimised results. *Ore Geology Reviews*, 112, 103015.

Kuyumjian, R. M., Araújo Filho, J. D. (2005). Depósitos e ocorrências de ouro no terreno arqueanopaleoproterozóico de Almas-Dianópolis (TO): evidências da importância metalogenética do evento Brasileiro. *Revista Brasileira de Geociências* 35(4), 611-614 (in portuguese).

Kuyumjian, R. M., Cruz, E. L. C. C., de Araújo Filho, J. O., Moura, M. A., Guimarães, E. M., & da Silva Pereira, K. M. (2012). Geologia e ocorrências de ouro do Terreno Granito-Greenstone do Tocantins, TO: síntese do conhecimento e parâmetros para exploração mineral. *Revista Brasileira de Geociências* 42(1), 213-228 (in portuguese).

Lary, D. J. (2010). Artificial intelligence in geoscience and remote sensing. In *Geoscience and Remote Sensing New Achievements*. IntechOpen.

Luiz, J. G., Silva, L. M. C. 1995. Geofísica de prospecção. Editora Cejup Ltda. Belém-PA. 311p.

Ma, Y., Minasny, B., & McBratney, A. (2021). Identifying soil provenance based on portable X-ray fluorescence measurements using similarity and inverse-mapping approaches – A case in the Lower Hunter Valley, Australia. *Geoderma Regional*, 25. <https://doi.org/10.1016/j.geodrs.2021.e00368>

MacLeod, I.N., & Ellis, R.G., 2013. Magnetic vector inversion, a simple approach to the challenge of varying direction of rock magnetization. *Australian Society of Exploration Geophysicists*, 1(4).

MacLeod, I.N., & Ellis, R.G., 2015. Quantitative magnetization vector inversion. *ASEG Extended Abstracts*, 2016(1), 1-6.

Mansur, E. T., & Ferreira Filho, C. F. (2016). Magmatic structure and geochemistry of the Luanga Mafic–Ultramafic Complex: Further constraints for the PGE-mineralized magmatism in Carajás, Brazil. *Lithos*, 266, 28-43.

Mansur, E. T., Ferreira Filho, C. F., & Oliveira, D. P. (2020). The Luanga deposit, Carajás Mineral Province, Brazil: Different styles of PGE mineralization hosted in a medium-size layered intrusion. *Ore Geology Reviews*, 118, 103340.

Mantovani, M. S., & de Brito Neves, B. B. (2009). The Paranapanema lithospheric block: its nature and role in the accretion of Gondwana. *Developments in Precambrian Geology*, 16, 257-272.

Marques, Gustavo Campos. Geologia dos Grupos Araí e Serra da Mesa e seu embasamento no sul do Tocantins. 2009. 122 f. Dissertação (Mestrado em Geologia) - Instituto de Geociências, Universidade de Brasília, Brasília, 2009.

Martins, E.S.F., & Vidotti, R.M., 2021. Another way of looking at an Alkaline Province. *Journal of Geodynamics*, 143, 101811. <https://doi.org/10.1016/j.jog.2020.101811>

- Martins-Ferreira, M. A. C., da Silva, H. C. D. M., & Rodrigues, E. N. Idade Mínima de Deposição do Greenstone Belt de Almas Indicada por Xenólitos Máficos e Sedimentares no Batólito Ribeirão das Areias. 2017. XV Simpósio de Geologia do Centro-Oeste. Goiânia. Resumos Expandidos (in portuguese).
- Martins-Ferreira, M.C., Dias, A.N.C., Chemale Jr, F., Campos, J.E.G., Seraine, M., Rodrigues, E.N., 2020. Multi-stage crustal accretion by magmatic flare-up and quiescence intervals in the western margin of the São Francisco Craton: U-Pb-Hf and geochemical constraints from the Almas Terrane. *Gondwana Research* 85, 32–54.
- McEnroe, S. A., Fabian, K., Robinson, P., Gaina, C., & Brown, L. L. (2009). Crustal magnetism, lamellar magnetism and rocks that remember. *Elements*, 5(4), 241-246.
- Mendes, T.L. (2023) Mapeamento da Assinatura de Ni-Cu (Co-Au-PGE) na Província Tocantins: o Exemplo da Intrusão Máfica-Ultramáfica Barra do Gameleira. (Dissertação de Mestrado).
- Michalski, R. S.; Carbonell, J. G.; Mitchell, T. M. Machine Learning: an Artificial Intelligence Approach. Germany: Springer Berlin Heidelberg, 1983. 572 p.
- Miller, H. G., & Singh, V. (1994). Potential field tilt—a new concept for location of potential field sources. *Journal of applied Geophysics*, 32(2-3), 213-217.
- Milligan, P. R., & Gunn, P. J. (1997). Enhancement and presentation of airborne geophysical data. *AGSO Journal of Australian Geology and Geophysics*, 17(2), 63-75.
- Minty, B. R. S. (1997). Fundamentals of airborne gamma-ray spectrometry. *AGSO Journal of Australian Geology and Geophysics*, 17, 39-50.iae
- Molugaram, K., Rao, G. S., Shah, A., & Davergave, N. (2017). *Statistical techniques for transportation engineering*. Butterworth-Heinemann.
- Nabighian, M. N. (1972). The analytic signal of two-dimensional magnetic bodies with polygonal cross-section: its properties and use for automated anomaly interpretation. *Geophysics*, 37(3), 507-517.
- Nabighian, M. N. (1984). Toward a three-dimensional automatic interpretation of potential field data via generalized Hilbert transforms: Fundamental relations. *Geophysics*, 49(6), 780- 786.
- Nabighian, M. N., Grauch, V. J. S., Hansen, R. O., LaFehr, T. R., Li, Y., Peirce, J. W., ... & Ruder, M. E. (2005). The historical development of the magnetic method in exploration. *Geophysics*, 70(6), 33ND-61ND.
- Naldrett, A. J. (1999). World-class Ni-Cu-PGE deposits: Key factors in their genesis. *Mineralium deposita*, 34, 227-240.
- Naldrett, A. J., & Naldrett, A. L. (2004). Magmatic sulfide deposits: Geology, geochemistry and exploration. Springer Science & Business Media. Oliveira, C.G. 2012. Projeto Natividade. Instituto de Geociências. Universidade de Brasília.
- Oliveira, C.G. 2012. Projeto Natividade. Instituto de Geociências. Universidade de Brasília.
- Oppong, F. B., & Agbedra, S. Y. (2016). Assessing univariate and multivariate normality. a guide for non-statisticians. *Mathematical theory and modeling*, 6(2), 26-33.

- Padilha, J. L. 1984. Prospecção de ouro na região nordeste de Goiás. Projeto Pindorama. DOCEGEO. In: SBG, Encontro Regional do Ouro de Goiás, 1, Goiânia. Anais, 78-95.
- Paris COP21 Agreement. 2015. Available at: <https://unfccc.int/process-and-meetings/the-paris-agreement/the-paris-agreement> (Last accessed, 13 Setembro 2023).
- Parsa, M., Lentz, D. R., & Walker, J. A. (2023). Predictive Modeling of Prospectivity for VHMS Mineral Deposits, Northeastern Bathurst Mining Camp, NB, Canada, Using an Ensemble Regularization Technique. *Natural Resources Research*, 32(1), 19-36.
- Paula-Santos, G. M., Babinski, M., Kuchenbecker, M., Caetano-Filho, S., Trindade, R. I., & Pedrosa-Soares, A. C. (2015). New evidence of an Ediacaran age for the Bambuí Group in southern São Francisco craton (eastern Brazil) from zircon U–Pb data and isotope chemostratigraphy. *Gondwana Research*, 28(2), 702-720.
- Pimentel M.M., Heaman L., Fuck R.A. 1991. Zircon and shpene Pb-U geochronology of Upper Proterozoic volcanic-arc rock units from southwestern Goiás, central Brazil. *Journal of South*.
- Pimentel, M.M., Ferreira Filho, C.F., Armstrong, R. A. (2004). SHRIMP U-Pb and Sm-Nd ages of the Niquelândia layered complex: Meso- (1,25Ga) and Neoproterozoic (0,79 Ga) extensional events in central Brazil. *Precambrian Research*, 132: 133-153.
- Pimentel, M.M., Rodrigues, J.B., Della Giustina, M.E.S., Junges, S.L., Matteini, M. (2011). The tectonic evolution of the Neoproterozoic Brasília Belt, Central Brazil, based on Shrimp and La-Icpms U-Pb sedimentary provenance data: a review. *Journal of South American Earth Sciences*, v. 31, p. 345-357.
- Pires, A. C. B. (1995). Identificação geofísica de áreas de alteração hidrotermal, Crixás-Guarinos, Goiás. *Brazilian Journal of Geology*, 25(1), 61-68.
- Pires, A.C.B. & Moraes R.A.V. (2006). New processing technologies applied do airborne geophysical data: impact on interpretation. 2º Simpósio Brasileiro de Exploração Mineral, SIMEXMIN, Ouro Preto, MG.
- Portela, A. C. P., Machetto, C. M. L., Santos, E. L. D., Meneguesso, G., Stein, J. H., Costa, L. A. M. D., ... & Silva, W. G. D. (1976). *Projeto leste do Tocantins/oeste do Rio São Francisco: relatório final*.
- Praxedes, I. F. (2015). Lineamento Transbrasiliiano, contribuição aerogeofísica, tectônica e geocronológica no setor nordeste da Faixa Brasília.
- Radford, D. D., Cracknell, M. J., Roach, M. J., & Cumming, G. V. (2018). Geological mapping in Western Tasmania using radar and random forests. *IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing*, 11(9), 3075-3087.
- Reeves, C. (2005). *Aeromagnetic surveys: principles, practice and interpretation* (Vol. 155, p. 2005). Washington (DC): Geosoft.
- Ribeiro, P. S. E., Frasca, A. A. S., Carneiro, J. D. S. M., Hattingh, K., Rezende, E. S. D., & Martins, F. R. (2022). Mapa geológico e de recursos minerais do Estado do Tocantins (em revisão).
- Ribeiro, V. B., Mantovani, M., & Louro, V. H. A. (2014). Aerogamaespectrometria e suas aplicações no mapeamento geológico. *Terra Didática*, 10(1), 29-51.

- Royston, J. P. (1982). An extension of Shapiro and Wilk's W test for normality to large samples. *Journal of the Royal Statistical Society: Series C (Applied Statistics)*, 31(2), 115-124.
- Russel, S., & Norvig, P. (2021). *Artificial Intelligence: A Modern Approach*, Global Edition. Foundations, 19, 1-1166.
- Saboia, A. M. (2021). *Geologia, geocronologia e geoquímica dos granitoides paleoproterozoicos do domínio Almas-Conceição do Tocantins, norte do orógeno Brasília: implicações magmáticas e geodinâmicas* (Doctoral dissertation).
- Saboia, A. M., de Oliveira, C. G., Dantas, E. L., Cordeiro, P., Scandolara, J. E., Rodrigues, J. B., & de Sousa, I. M. C. (2020a). The Siderian crust (2.47–2.3 Ga) of the Goiás Massif and its role as a building block of the São Francisco paleocontinent. *Precambrian Research*, 350, 105901.
- Saboia, A. M., de Oliveira, C. G., Dantas, E. L., Scandolara, J. E., Cordeiro, P., Rodrigues, J. B., & de Sousa, I. M. C. (2020b). The 2.26 to 2.18 Ga Arc-Related Magmatism of the Almas-Conceição do Tocantins Domain: An Early Stage of the São Francisco Paleocontinent Assembly in Central Brazil. *Journal of South American Earth Sciences*, 104, 102757.
- Saboia, A.M., Meneghini, P.F.V.B., (2019). *Geologia e Recursos Minerais da Folha Dianópolis - SC.23-Y-C: Escala 1:250.000, Projeto Sudeste do Tocantins, estado do Tocantins – Goiânia: CPRM*, 193 p. (in portuguese).
- Saboia, A.M., Meneghini, P.F.V.B., Gonçalves, G. D. F., Lacerda Filho, J.V.D., (2014). *Carta geológica: folha Dianópolis-SC. 23-YC.* (in portuguese).
- Saunders, D. F., Burson, K. R., Branch, J. F., & Thompson, C. K. (1993). Relation of thorium-normalized surface and aerial radiometric data to subsurface petroleum accumulations. *Geophysics*, 58(10), 1417-1427.
- Saunders, D. F., Terry, S. A., & Thompson, C. K. (1987). Test of national uranium resource evaluation gamma-ray spectral data in petroleum reconnaissance. *Geophysics*, 52(11), 1547-1556.
- Schnitzler, N., Ross, P. S., & Gloaguen, E. (2019). Using machine learning to estimate a key missing geochemical variable in mining exploration: Application of the Random Forest algorithm to multi-sensor core logging data. *Journal of Geochemical Exploration*, 205, 106344.
- Schön, J. (2011). *Physical properties of rocks: A workbook* (Vol. 8). Elsevier.
- Shapiro, S. S., & Wilk, M. B., (1965). An analysis of variance test for normality (complete samples). *Biometrika*, 52(3/4), 591-611.
- Silva, J. M. (2009). *O Complexo Máfico-ultramáfico Acamadado de Americano do Brasil e sua mineralização de Ni-Cu-Co*.
- Skirrow, R. G., Huston, D. L., Mernagh, T. P., Thorne, J. P., Duffer, H., & Senior, A. (2013). Critical commodities for a high-tech world: Australia's potential to supply global demand (p. 126). Canberra: Geoscience Australia.

Sousa, I.M.C., Giustina, M.E.S.D., Oliveira, C.G., (2016). Crustal evolution of the northern Brasília Belt basement, central Brazil: A Rhyacian orogeny coeval with a pre-Rodinia supercontinent assembly. *Precambrian Research*, 273,129-150.

Sutton, R. S., & Barton, A. G. (1998). Introduction to RL. Reinforcement Learning: an introduction.

Telford W.M., Geldart L.P., Sheriff R.E. (1990). *Applied Geophysics*. 2ed. Cambridge University Press. 770 p. <https://doi.org/10.1017/cbo9781139167932>.

Verduzco, B., Fairhead, J. D., Green, C. M., MacKenzie, C. (2004). New insights into magnetic derivatives for structural mapping. *The Leading Edge* 23, 116–119.

Wilford, J., & Minty, B. (2006). The use of airborne gamma-ray imagery for mapping soils and understanding landscape processes. *Developments in soil science*, 31, 207-610.

Yandjimain, J., Ndougsa-Mbarga, T., Bi-Alou, M. B., & Meying, A. (2018). Aeromagnetic Data Modeling for Geological and Structural Mappings over the DJADOM-ETA Area, in the Southeastern Cameroon. *International Journal of Geosciences*, 9(06), 354.

Yu, L., Porwal, A., Holden, E. J., & Dentith, M. C. (2012). Towards automatic lithological classification from remote sensing data using support vector machines. *Computers & Geosciences*, 45, 229-239.

Zhang, N.; Zhou, K.; Li, D. (2018). Back-propagation neural network and support vector machines for gold mineral prospectivity mapping in the Hatu region, Xinjiang, China. *Earth Science Informatics* 2018 11:4, v. 11, n. 4, p. 553–566.