

Avaliação de Dispositivos Raspberry PI Para Seleção e Classificação de Bandas de Imagens Hiperespectrais usando Algoritmos Genéticos e KNN

Luiz André Marques Furlan dos Santos, Diego Saqui (Orientador)

Instituto Federal de Educação, Ciência e Tecnologia, Corumbá –MS

RESUMO

Esta pesquisa apresenta uma avaliação da eficiência do uso do Raspberry PI nos quesitos de execução do processamento e diminuição dos custos computacionais para a seleção de bandas e classificação de imagens hiperespectrais utilizando algoritmos de inteligência artificial. Para tanto, foi utilizado um algoritmo genético (AG) e o classificador *K-nearest neighbor* (KNN), executado em 4 equipamentos diferentes. Os resultados foram observados para fazer um comparativo e, assim poder compreender se é ou não viável o uso do Raspberry PI. Adicionalmente, além da avaliação do dispositivo Raspberry PI, também foi observada a taxa de acerto da classificação das bandas hiperespectrais, usando o classificador KNN. Previamente, foram realizadas pesquisas bibliográficas diversas, para compreender a melhor forma de adaptar um algoritmo que fosse capaz de ser executado da forma necessária, em todos os dispositivos que seriam utilizados na pesquisa. Após esse período, o mesmo foi posto em execução em cada dispositivo em separado, com resultados sendo monitorados diariamente. Ao final, os resultados obtidos demonstraram que, de um modo geral, o uso do equipamento elegido (Raspberry PI) é viável para o trabalho proposto. A execução do algoritmo se deu de forma satisfatória e, embora o tempo de execução tenha sido maior do que nos outros equipamentos, há de se considerar que o custo extremamente baixo torna viável seu uso.

Palavras-chave: Raspberry PI, Algoritmo Genético, KNN.

1 INTRODUÇÃO

Atualmente, a preocupação com facilitar as técnicas de monitoramento aéreo tem se tornado um grande desafio tecnológico. Facilitar o acesso ao conhecimento e armazenamento de informações que possam ser utilizadas em diversas situações, pode ser de imensa relevância, principalmente para os processos de agricultura.

Nesse contexto, as imagens hiperespectrais (IH) são úteis porque possibilitam uma quantidade de aplicações muito grande em relação a outros tipos de imagens, podendo ser capturadas por técnicas de sensoriamento remoto. Essas imagens possuem maior número de bandas espectrais em relação a imagens multiespectrais e com bandas tradicionais, ou seja, *red green blue* (RGB), que possibilitam maior desempenho no processo de segmentação.

No ramo do agronegócio, as IHS, perfazem uma dinâmica interessante, pois possibilitam a

identificação de áreas específicas, dando acesso para que, através delas, se torne mais fácil identificar as necessidades no manejo da área.

Para captura dessas imagens, cada vez mais se faz uso de veículos aéreos não tripulados, mais popularmente denominados drones. Em seu uso, os drones são destinados a sobrevoar áreas de interesse para monitoramento aéreo, através da captura de imagens hiperespectrais.

Formaggio *et al.* (2017) afirmam que o sensoriamento remoto pode trazer importantes contribuições no fornecimento de dados precisos para o setor agrícola.

Como bem demonstrado por Belcore *et al.*, (2019), as vantagens do uso de drones são inúmeras, pois eles podem voar a altitudes baixas e alcançam áreas remotas, além de poder ser equipados com uma infinidade de sensores.

Breunig *et al.* (2019) definem que “ao integrar um potencial de alta resolução espacial com uma grande

repetitividade, os drones podem acompanhar de forma adequada a fenologia de culturas agrícolas, demandas dos processos de manejo da lavoura e qualificar a tomada de decisões”.

Assim, o uso do monitoramento aéreo através de drones que capturam IHS tem uma relevância dentro do planejamento da agricultura, e facilitar o acesso às tecnologias de forma mais barata é o objetivo do presente trabalho.

O presente trabalho foi desenvolvido com o objetivo de buscar uma forma eficiente de melhorar a acurácia e reduzir o custo para a tarefa de seleção de bandas de IHS e posteriormente, a classificação dessas imagens usando dispositivos Raspberry PI de baixo custo. Assim, iniciou-se conceituando a IHS, demonstrando qual sua aplicação no plano prático, para uma melhor compreensão de sua aplicabilidade. Em seguida, foram enumeradas as formas de extração de bandas e explicada a metodologia utilizada na presente pesquisa, para, ao final, demonstrar os resultados obtidos.

2 REFERENCIAL TEÓRICO

Muitas são as formas de utilização de imagens na atualidade. Entre as imagens de uso complexo e abrangente, encontram-se as chamadas imagens hiperespectrais (IHS), que vêm ganhando destaque por sua ampla utilização e flexibilidade;

Para Fauvel et al. (2012), os recentes avanços tecnológicos na área de sensoriamento remoto hiperespectral, acabam por permitir a simultânea aquisição de centenas de espectros para cada pixel de imagem. Ainda no pensamento desse autor, a informação espectral dessa forma detalhada majora as possibilidades de materiais de interesse mais específicos, vez que, além disso, possui na resolução espacial dos sensores uma maior facilidade de análise de pequenas estruturas espaciais na imagem.

Os autores ainda explicam que, dentro do domínio espectral, os pixels são representados por vetores, e nesses, cada componente é uma medida que corresponde a determinado comprimento de onda específico. Assim, o tamanho do vetor é igual ao número de bandas espectrais que o sensor consegue coletar. Nas IHS, muitas centenas de bandas espectrais da mesma cena são disponíveis, enquanto

que nas imagens multiespectrais, no máximo dez bandas são fornecidas.

A crescente necessidade de expandir tecnologicamente as possibilidades de monitoramento, coincide com a necessidade de monitoramento do meio ambiente numa escala global e também da obtenção de melhores informações acerca dos recursos naturais renováveis e não renováveis. Dessa necessidade, surgiu no Brasil a presença do ensino do sensoriamento remoto em diversos cursos universitários, de forma ampla, abrangendo desde as “ciências naturais da Terra, à computação, urbanismo, engenharia civil, geotecnia, cartografia, ordenamento territorial, agricultura, geologia, defesa civil e muitas outras” (MENESES, 2012, p.4).

Ainda nesse pensamento, Meneses (2012) afirma que a multidisciplinaridade do sensoriamento permite que sua utilização seja facilmente incorporada nas mais variadas áreas. Entretanto, é necessário compreender como ler os dados digitais das imagens e também como fazer seu processamento. Atualmente, a grande diversidade de programas de processamentos de imagens de sensoriamento remoto busca facilitar a obtenção de dados para os mais variados usuários. O uso de IHS captadas por sensoriamento remoto é uma das ferramentas de grande utilização nos dias atuais, sendo certo que é uma das formas mais precisas de monitoramento existentes, dada a sua capacidade de captação de características por imagem.

Num contexto de atualidade, as tecnologias vêm sendo amplamente utilizadas em benefício humano. Assim, pode-se observar as chamadas IHS, que são compostas por um grande número de bandas espectrais. Essas bandas presentes nas IHS são úteis por possibilitar uma quantidade de aplicações muito grande em relação a outros tipos de imagens.

IHS são úteis porque possibilitam uma quantidade de aplicações muito grande em relação a outros tipos de imagens (multiespectrais ou RGB). Podem ser capturadas por técnicas de sensoriamento remoto que permitem diversos tipos de aplicações, como a produção de mapas temáticos de áreas agrícolas. Essas imagens possuem bandas adicionais em relação a imagens multiespectrais e RGB, que possibilitam maior desempenho no processo de

segmentação. Assim, a extração de características é um aspecto que propõe realizar uma modificação no espaço de características para um novo espaço de dimensão reduzida. Entretanto, os dados subjetivos podem ser distorcidos e a informação física se perder no processo de transformação (ZHANG; GONG; CHAN, 2015).

Dessa forma, na presente pesquisa, imagens já captadas por meio de sensoriamento hiperespectral foram analisadas, para que seja possível compreender uma melhor forma de processamento dos dados extraídos, com um menor custo.

Apesar das técnicas de segmentação, dos mapas temáticos e IHS proporcionarem inúmeras vantagens, ainda subsiste um desafio que se relaciona ao volume de dados e a dimensão dessas imagens e, por isso, alguns problemas são bastante recorrentes, tais como o armazenamento, transmissão e tempo de processamento de dados de IHS (SOHAIB *et al*, 2013); informações redundantes, causadas pela alta correlação entre bandas, especialmente as adjacentes, que podem influenciar no desempenho de classificadores tradicionais (DATTA *et al*, 2012); fenômeno de Hughes (maldição da dimensionalidade), apresentado por Hughes (HUGHES, 1982). Além desse aspecto, ainda é necessário observar que os custos computacionais para esse tipo de execução, torna o uso dessa técnica um tanto inacessível. Portanto, a presente pesquisa se presta a alcançar uma forma prática, eficiente e mais econômica de utilização das IHS.

A seleção de bandas consiste em selecionar quais bandas capturadas pelas imagens de sensoriamento possuem valores não duplicados e não nulos, fazendo a separação de forma a salvar os dados realmente relevantes. Patra, Modi e Buzzzone (2015) apresentam como vantagens do uso dessa abordagem alguns aspectos, como a possibilidade de validação conceitual das bandas selecionadas e, assim, da informação usada pelo classificador; a possibilidade de utilização dos resultados da seleção como ferramenta de mineração de dados que possibilita alcançar informação física (bandas espectrais) baseado em classes discriminadas; seleção de subconjunto de bandas que resulta na possibilidade de definição de um sistema em que bandas irrelevantes não sejam armazenadas.

Na pesquisa realizada, foram utilizados ainda os chamados Algoritmos Genéticos (AGs) que, para Rivero et al. (2010), são métodos considerados adaptativos, baseados na reprodução sexual e na sobrevivência do mais apto. De forma mais específica, pode-se considerar que são algoritmos de busca baseados na seleção natural (genética natural). AG combinam a sobrevivência do mais apto entre as estruturas com uma troca de informações mais estruturada, mesmo que de forma aleatória, para dessa forma constituir um algoritmo de busca que tenha alguma semelhança com as buscas humanas (GOLDBERG, 1989 APUD RIVERO ET AL., 2010).

Um método de classificação tradicional na área de sensoriamento remoto e processamento de imagens aéreas é o k-vizinhos mais próximos (KNN), que para Mourão (2013), consiste em um método de determinação sem a utilização de parâmetros, que classifica determinada imagem através do cálculo de uma métrica (distâncias vetoriais – distância euclidiana) entre a imagem a ser classificada, ainda desconhecida e cada imagem do conjunto de treinamento - conhecidas e previamente rotuladas como normais ou anormais).

Na presente pesquisa, o algoritmo desenvolvido foi genético incorporando um classificador KNN no processo de seleção de bandas, caracterizando a abordagem como *Wrapper*.

Conforme Almeida (2018), a abordagem *Wrapper* funciona de forma que o algoritmo de classificação “é executado para cada subconjunto e a avaliação geralmente é feita em termos da acurácia preditiva retornada pelo algoritmo” (ALMEIDA, 2018). Dessa forma, o subconjunto que apresentar o melhor desempenho no aprendizado é definido como o melhor subconjunto.

Na presente pesquisa, o método *wrapper* utilizado se baseia na divisão 3 (três) subconjuntos, sendo 2 (dois) para treinamento e 1 (um) para testes.

Para Richard e Jia (2005), classificação supervisionada consiste em um procedimento de análise de dados de imagens de sensoriamento remoto. Nesse modelo de procedimento, considera-se a existência de pixels rotulados de uma imagem chamada *Ground Truth* (GT). Algumas etapas desses algoritmos de classificação supervisionada para

imagens de sensoriamento remoto podem ser elencadas como:

1-) decidir um conjunto de coberturas do solo da imagem que serão segmentadas (água, areia e vegetação, etc.);

2-) escolha de elementos representativos (dados de treinamento para cada cobertura do solo desejada, por exemplo);

3-) utilização dos dados de treinamento para treinar um algoritmo de classificação;

4-) utilização do classificador treinado com os dados de treinamento para classificação de novos dados;

5-) geração de mapas temáticos que resumem os resultados da classificação.

Como explica Vujovic (2014), o Raspberry Pi é uma placa de computador pequena, poderosa, barata e orientada para a educação, lançada em 2012. Do tamanho de um cartão de crédito, se apresenta com bom desempenho e de preço acessível, é a plataforma perfeita para a interface com muitos dispositivos.

O Raspberry Pi contém uma placa gráfica, memória RAM e várias interfaces e conectores para dispositivos externos. Este dispositivo opera da mesma forma que um computador convencional, exigindo teclado, mouse, monitor e uma fonte de alimentação. Assim como os computadores convencionais, o Raspberry Pi também usa um sistema operacional, normalmente baseado em Linux.

O Linux, como um sistema operacional gratuito e de código aberto, é uma excelente combinação para o Raspberry Pi, especialmente por contribuir para manter o baixo custo do equipamento (Vujovic, 2014).

3 MATERIAIS E MÉTODOS

Na presente pesquisa foi utilizado um Algoritmo Genético (AG) em conjunto com o classificador KNN. O AG foi desenvolvido na linguagem Python.

A população inicial é composta pela base de dados gerada através da imagem na sua forma original, sem qualquer tratamento. No caso da pesquisa realizada, a base de dados utilizada foi extraída da Indian Pines, imagem hiperespectral, capturada através de sensoriamento remoto, no estado de Indiana (Estados Unidos). Essa imagem contém 220 bandas espectrais e 16 classes. As classes correspondem a diversos tipos de plantio (milho, alfafa, etc.). Essa imagem está disponível para *download*¹ de forma gratuita no site da Faculdade de Engenharia da Universidade de Purdue. Essa Faculdade está localizada no estado de Indiana, Estados Unidos.

Na figura 1 é demonstrado um exemplo do AG utilizado na presente pesquisa. Na primeira etapa (1), é apresentada a população inicial definida, na segunda etapa (2), o algoritmo promove uma seleção dos indivíduos mais aptos, utilizando o classificador KNN para validar a seleção, foi considerado $k=1$ na presente pesquisa, após a seleção o algoritmo promove o cruzamento(reprodução) entre os indivíduos (3). Essa reprodução promove a transmissão de características (4) de uma geração para outra, o que resulta em mutações (5) dessas características, conforme as gerações forem evoluindo. Com a mutação, uma nova população (6) é constituída, voltando ao início do ciclo, se repetindo quantas vezes o usuário determinar (7), até que finalize sua execução (8).

A escolha do equipamento utilizado foi pensada para atender o maior número possível de potenciais usuários de sensoriamento remoto com uso de imagens hiperespectrais. Para tanto, foi selecionado um equipamento de maior popularidade, levando-se em consideração o custo de aquisição e a facilidade de acesso e operação do mesmo. Como explicado no capítulo anterior, o Raspberry PI funciona como um computador convencional e isso viabiliza a execução da pesquisa, por se tratar de um equipamento de baixo custo que possui todas as propriedades necessárias para executar o algoritmo.

1

http://www.ehu.eus/ccwintco/index.php/Hyperspectral_Remote_Sensing_Scenes

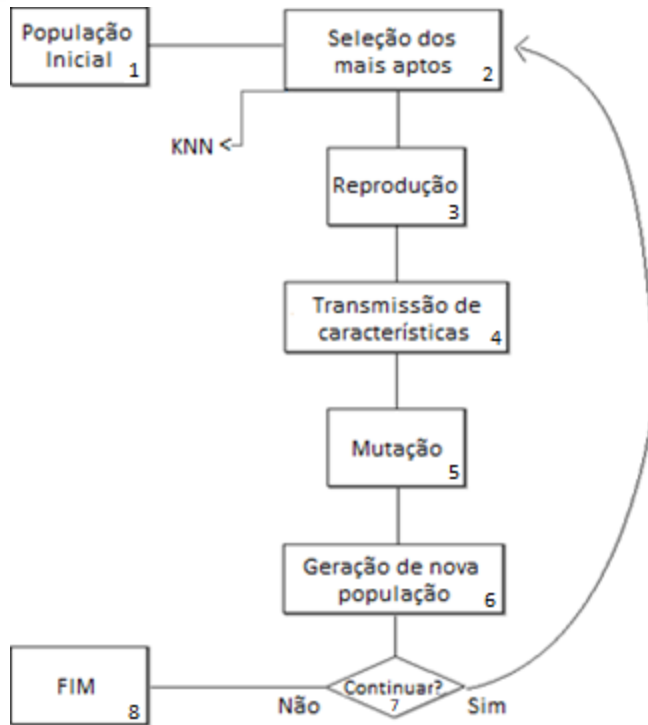


Figura 1: Fluxograma de funcionamento do algoritmo gen tico. A numera  o corresponde as etapas.

Nesse sentido, o equipamento escolhido para a pesquisa foi o Raspberry PI em duas vers es: (Zero e 3B), testando sua efici ncia em menor custo do projeto como um todo, levando em considera  o o custo do equipamento, vantagens gerais, como fontes alternativas de alimenta  o, j  que os dispositivos podem ser alimentados por uma bateria de celular, em caso de queda de energia, permanecer o ligados por v rias horas, se mantendo apenas com a energia da bateria. Ao final da pesquisa, o resultado foi comparado ao obtido com o uso de dois *notebooks* convencionais.

4 RESULTADOS E DISCUSS O

4.1 AVALIA  O DOS DISPOSITIVOS

Ap s executadas 200 (duzentas) gera  es do Algoritmo Gen tico (AG) em cada dispositivo, o tempo de execu  o no Raspberry PI 3 foi de 67h20min (sessenta e sete horas e vinte minutos),

enquanto que no Raspberry PI zero foi de 312h30min (trezentas e doze horas e trinta minutos) e para o *notebook* SAMSUNG RV 410 foi de 15h (quinze horas) e para o *notebook* POSITIVO MOTION Q232A foi de 38h (trinta e oito horas).

A tabela 1 faz uma breve compara  o de *hardware* entre os quatro dispositivos citados.

Marca	Samsung	Positivo	Raspberry PI	Raspbery PI
Modelo	RV410	Q232A	3B	Zero
Mem�ria RAM	4GB	2GB	1GB	512MB
Ddr (RAM)	3	3	2	2
Velocidade (RAM)	1333mhz	1066mhz	900MHz	900MHz
Processador	2x 2.3GHz	4x 1.92GHz	4x 1.2GHz	1x 1GHz
Velocidade (ROM)	300MB/s	280MB/s	25MB/s	12.5MB/s
Sistema Operacional	Deepin Linux	Deepin Linux	Raspbian	Raspbian
Custo	R\$900,00	R\$ 1000,00	R\$150,00	R\$25,00

Tabela 1. Comparativo de *hardware* entre os dispositivos

Com o decorrer das gera  es do algoritmo gen tico, a taxa de acerto do classificador tende a aumentar. Conforme as gera  es passam, maior a porcentagem de acerto. Pode-se concluir que o uso do Raspberry PI realmente diminui o custo computacional na execu  o da tarefa proposta (processamento de imagens hiperespectrais), entretanto, o tempo de processamento   mais longo do que observado nos equipamentos utilizados para compara  o (notebooks), isso se d  pela limita  o de *hardware* do Raspberry PI. Mesmo assim, pode-se considerar vi vel a utiliza  o do Raspberry PI como alternativa de diminui  o de custos com efici ncia na execu  o do processamento.

Os gr ficos nas figuras 2,3,4 e 5 representam o aumento da taxa de acerto do classificador para cada um dos dispositivos ao longo das gera  es.

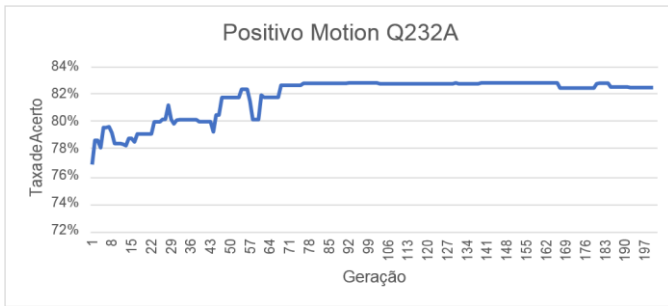


Figura 2: Gráfico da taxa de acerto do classificador no Positivo Motion Q232A.

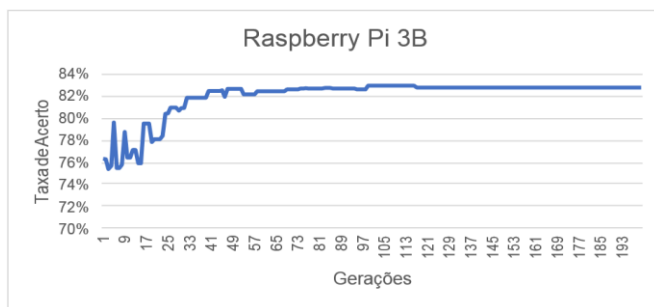


Figura 3: Gráfico da taxa de acerto do classificador no Raspberry PI 3B.

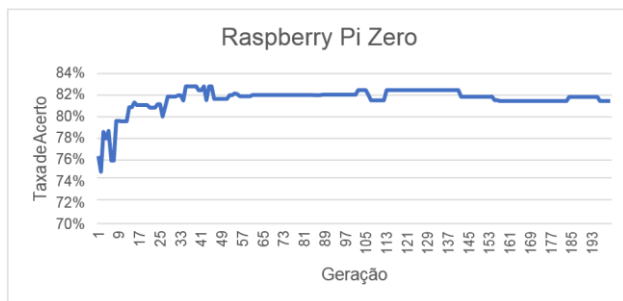


Figura 4: Gráfico da taxa de acerto do classificador no Raspberry PI Zero.

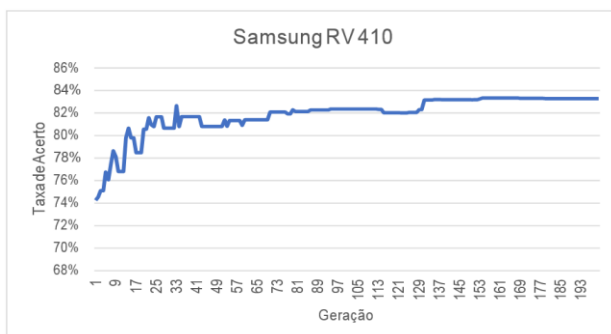


Figura 5: Gráfico da taxa de acerto do classificador no Samsung RV 410.

4.2 AVALIAÇÃO DOS CLASSIFICADORES

A base de dados utilizada foi dividida em dois grandes grupos, sendo eles para treinamento e testes.

O AG retornou o conjunto de bandas espectrais por ele selecionado. Esse conjunto foi utilizado em um processo de classificação em cima da base de dados da imagem hiperespectral utilizada, utilizando os grupos de treinamento e testes, considerando o número K de vizinhos mais próximos sendo este igual a 1, 3, 5, 7 e 9, para cada um dos dispositivos testados. A tabela a seguir demonstra a taxa de acerto do classificador KNN após o processo anteriormente descrito.

Dispositivo	K=1	K=3	K=5	K=7	K=9
Rv410	55.4	57.6	60.3	62.0	61.9
Q232A	57.8	61.6	63.4	64.8	65.3
Raspberry 3B	59.9	62.2	65.3	66.8	67.7
Raspberry 0	58.3	61.7	62.7	62.9	62.8

Tabela 2. Comparativo de acurácia entre os dispositivos utilizados. Os valores apresentados representam a porcentagem de acerto.

5 CONCLUSÃO

Após a observação da execução do algoritmo nos diversos dispositivos elegidos, observou-se que os equipamentos Raspberry PI, nos dois modelos utilizados, processaram os dados de forma satisfatória.

Dentro do que se pretendia ao início da pesquisa, ficou mais do que evidente a eficiência do Raspberry PI, considerando seu baixo custo e a execução satisfatória do algoritmo, com taxa de acerto satisfatória.

O algoritmo genético mostrou-se eficiente no aumento da taxa de acerto da classificação das bandas das IHS ao longo das gerações. Portanto, o uso do método utilizado pode, em muito, contribuir para uma execução satisfatória muito mais barata do que as que habitualmente vêm sendo utilizadas, como demonstrado na Tabela 1, vez que o Raspberry

PI Zero custa em torno R\$ 25,00 (vinte e cinco reais – importado com valor estimado em conversão direta), o Raspberry PI 3B custa em média R\$ 150,00 (cento e cinquenta reais – importado com valor estimado em conversão direta) e os notebooks utilizados têm preço médio de R\$ 1000,00 (mil reais).

Dessa forma, conclui-se que o objetivo da presente pesquisa restou alcançado, ficando demonstrada a possibilidade de uso do Raspberry PI de forma eficiente, diminuindo assim o custo de processamento das IHS e facilitando o acesso às tecnologias de monitoramento remoto.

AGRADECIMENTOS

Agradeço ao Instituto Federal de Mato Grosso do Sul, por possibilitar a realização dessa frente de pesquisa e pelo incentivo à ciência.

A minha mãe, Danielle, por me apoiar nos momentos difíceis.

A Gabriellen Vitória, pelo café, pelos memes e pelo apoio incondicional em todos os momentos de desespero.

REFERÊNCIAS

BELCORE, Elena, et al. (2019) "Raspberry PI 3 multispectral low-cost sensor for UAV based remote sensing. Case study in south-west Niger." Int. Arch. Photogramm. Remote Sens. Spat. Inf. Sci

BREUNIG, Fábio Marcelo, et al. (2019) "Drones e Ciência"

DATTA, A.; GHOSH, S.; GHOSH, A. (2012) "Clustering based band selection for hyperspectral images", In the International Conference on Communications, Devices and Intelligent Systems (CODIS)

FAUVEL, Mathieu et al. (2012) "Advances in spectral-spatial classification of hyperspectral images", Proceedings of the IEEE, v. 101, n. 3

FORMAGGIO, Antonio Roberto et al. (2017) "Sensoriamento remoto em agricultura". Oficina de Textos

GONG, M.; Zhang, M.; Yuan, Y. (2015)

"Unsupervised Band Selection Based Evolutionary Multiobjective Optimization for Hyperspectral Images", IEEE Transactions on Geoscience and Remote Sensing

HUGHES, G. F. (1982) "On the mean accuracy of statistical pattern recognizer", IEEE Trans. Inform. Theory, vol. IT-28

LEE, C.; Landgrebe, D. A. (1993) "Analyzing High-Dimensional Multispectral Data", IEEE Transactions on Geoscience and Remote Sensing, 31(4)

MENESES, Paulo Roberto et al. (2012) "Introdução ao processamento de imagens de sensoriamento remoto", Universidade de Brasília, Brasília

MOURAO, Helmer As; De Oliveira, Horácio AB Fernandes. (2013) "Localização de Dispositivos Móveis em Redes WiFi usando Variação da Potência de Transmissão e KNN"

PATRA, S.; Modi, P.; Bruzzone, L. (2015) "Hyperspectral Band Selection Based on Rough Set", IEEE Transactions on Geoscience and Remote Sensing

RICHARDS, J. A.; Jia, X. (2005) "Remote Sensing Digital Image Analysis. An Introduction", (4 ed.). Springer

RIVERO, Marcos Gestal Daniel et al. (2010) "Introducción a los algoritmos genéticos y la programación genética", Universidad de Coruña

SOHAIB, M.; Ulhaq, I.; Mushtaq, Q. (2013) "Dimensional reduction of hyperspectral image data using band clustering and selection based on statistical characteristics of band images", In the International Journal of Computer and Communication Engineering