

ESTIMATIVA DE UMIDADE DO SOLO POR MEIO DE APRENDIZADO DE MÁQUINA USANDO IMAGENS DE VEICULO AÉREO NÃO TRIPULADO (VANT)

ANDERSON LUIZ DOS SANTOS SAFRE¹; CAIO NASCIMENTO FERNANDES²;
JOÃO CARLOS CURY SAAD³

¹Aluno de Doutorado em Irrigação e Drenagem, Departamento de Engenharia Rural, UNESP-Faculdade de Ciências Agrônomicas, R. José Barbosa de Barros, 1780, CEP 18610-034, Botucatu-SP, Brasil. E-mail: andersonsafre@gmail.com

²Aluno de Mestrado em Irrigação e Drenagem, Departamento de Engenharia Rural, UNESP-Faculdade de Ciências Agrônomicas, R. José Barbosa de Barros, 1780, CEP 18610-034, Botucatu-SP, Brasil. E-mail: caionfernandes@hotmail.com

³Professor Titular, Departamento de Engenharia Rural, UNESP-Faculdade de Ciências Agrônomicas, R. José Barbosa de Barros, 1780, CEP 18610-034, Botucatu-SP, Brasil. E-mail: joaosaad@fca.unesp.br

1 RESUMO

A umidade do solo é um parâmetro importante para o cálculo da lâmina e manejo da irrigação, pois está diretamente relacionada ao conteúdo de água no solo. Técnicas de sensoriamento remoto aliadas a modelos estatísticos podem ser usadas para estimar a variabilidade espacial da umidade do solo, extrapolando medidas pontuais. O objetivo desse estudo foi determinar a umidade do solo por meio de algoritmos de *machine learning* (aprendizado de máquina) como *Support Vector Regression* (SVR), *Random Forests* (RF) e *Artificial Neural Networks* (ANN). Utilizou-se imagens multiespectrais de alta resolução adquiridas por meio de um Veículo Aéreo Não Tripulado (VANT) em uma área de feijão irrigado na Fazenda Experimental Lageado da Unesp, em Botucatu, SP, Brasil. Adotou-se como dados de entrada nos modelos, as refletâncias nas bandas do verde, vermelho, infravermelho próximo e o NDVI. Todos os algoritmos tiveram performance adequada, porém o modelo que melhor estimou a umidade do solo foi o SVR, com erro médio quadrático (RMSE) de 0,46 vol. % e coeficiente de determinação (R²) de 0,71.

Palavras-chave: umidade do solo, aprendizado de máquinas, VANT, redes neurais.

SAFRE, A. L. S.; FERNANDES, C. N.; SAAD, J. C. C. SOIL MOISTURE ESTIMATION THROUGH MACHINE LEARNING USING UNMANNED AERIAL VEHICLE (UAV) IMAGES

2 ABSTRACT

The soil moisture is an important parameter for the calculation of water depth and irrigation management since it is directly related to the soil water content. Remote sensing techniques combined with statistical models can be used to estimate the spatial variability of soil moisture, extrapolating point measurements. The objective of this study was to determine the soil moisture through machine learning algorithms such as Support Vector Regression (SVR), Random Forests (RF), and Artificial Neural Networks (ANN). High resolution multispectral images obtained by an Unmanned Aerial Vehicle (UAV) in an irrigated bean area at the Experimental Lageado Farm at Unesp in Botucatu, SP, Brazil, were used. The reflectances in

the Green, Red and Near Infrared bands along with the NDVI vegetation index were used as inputs for the models. All the algorithms performed well; however, the model that best fitted the data was the SVR, with mean square error (RMSE) of 0.46% of the estimated soil moisture and determination coefficient (R^2) of 0.71.

Keywords: soil moisture, machine learning, UAV, artificial neural networks.

3 INTRODUÇÃO

A disponibilidade dos recursos hídricos e energéticos para a agricultura é cada vez mais limitada devido às mudanças climáticas e à demanda crescente por alimentos. Os sistemas agrícolas, principalmente os irrigados, necessitam de técnicas e manejos que proporcionem altos níveis de eficiência para que haja a preservação dos recursos naturais ao mesmo tempo em que se configuram sistemas de produção agrícolas mais competitivos. Deste modo, o monitoramento da umidade do solo é uma das práticas fundamentais para se realizar um manejo eficiente e racional dos sistemas agrícolas irrigados, visto que a aplicação dessa técnica permite um melhor desenvolvimento das culturas e uma utilização otimizada tanto dos recursos hídricos como dos recursos e energéticos (HUISMAN et al., 2003; DUKES; ZOTARELLI; MORGAN, 2010; MONTESANO et al., 2015; BRITO et al., 2009; FREITAS et al., 2012; BRITO et al., 2014).

A obtenção da umidade do solo pode ser feita por meio de métodos diretos ou indiretos. A determinação da umidade pelo método direto é feita por meio da diferença entre a massa da amostra de solo em seu estado inicial e após essa ser seca. Resistência do solo à passagem de corrente elétrica, sonda de nêutrons, sensores capacitivos e tensão de água no solo, são exemplos de métodos indiretos de determinação da umidade do solo (TEIXEIRA; COELHO, 2005; DOBRIYAL et al., 2012). O uso do sensoriamento remoto também é um método indireto, pois por meio

de imagens captadas por câmeras acopladas à satélites ou veículos aéreos não tripulados (VANTs) é possível relacionar a umidade do solo com a variação da radiação eletromagnética refletida, utilizando para tal, métodos estatísticos.

A medição da tensão de água no solo com equipamentos denominados tensiômetros é uma das técnicas mais conhecidas de determinação indireta da umidade. Além de ter menor custo, quando comparada às demais técnicas, apresenta fácil manipulação, alta precisão e possibilidade de automação do sistema de leitura (ARRUDA et al., 2017) e do próprio sistema de irrigação (MONTESANO et al., 2015). A quantificação do teor de água no solo de forma precisa é necessária para auxiliar a tomada de decisão do produtor de quando e quanto irrigar e assim, ter um manejo eficiente e racional (THALHEIMER, 2013), fornecendo a quantidade de água adequada ao desenvolvimento das culturas de forma plena, resultando em ganhos de produtividade.

No desenvolvimento inicial do tensiômetro, o elemento químico mercúrio foi utilizado como escala de medida, sendo implementado por Livingston em 1908 (OR, 2001), apresentando a capacidade de medir o potencial mátrico do solo (LIBARDI, 2005). O contato do solo com a capsula porosa do tensiômetro proporciona um equilíbrio do meio. Conforme a umidade do solo diminui, a água presente no tensiômetro é sorvida pela tensão matricial gerada, reduzindo a pressão interna do sistema e ocasionando um aumento da tensão no tensiômetro. A leitura da tensão é comparada

com a curva de retenção de água característica do solo e assim, a umidade do solo pode ser determinada (CAMARGO; GROHMANN; CAMARGO, 1982).

A agricultura de precisão possibilita aos produtores monitorar as condições específicas de cada local, possibilitando um manejo extremamente eficiente, justamente por apresentar como princípio principal, o uso de parâmetros específicos da propriedade agrícola. Dentro da agricultura de precisão, podemos destacar os sistemas de informações espaciais, como o sistema de informação geográfica (SIG), ferramentas de sensoriamento remoto e os sistemas de posicionamento global.

Pode-se definir o sensoriamento remoto como o método de aquisição de informações de um determinado fenômeno ou comportamento da superfície terrestre sem existir o contato físico (JENSEN, 2009). Esse método permite a identificação de características de alvos específicos a partir da interação com a radiação eletromagnética (REM) (ROSA, 2009). Para isso, são utilizados sensores acoplados em satélites, aeronaves, VANTs e outras plataformas.

A REM é a energia que se desloca na velocidade da luz, seja na forma de partículas ou ondas eletromagnéticas, não sendo necessário um meio físico para sua propagação (ROSA, 2009). Composto a REM, está o espectro eletromagnético que envolve o intervalo de todos os comprimentos de onda, desde os raios gama, até as ondas de rádio (NOVO, 2008). Ao se propagar pelo espaço, o fluxo de radiação eletromagnética pode ou não interagir com objetos ou a superfície e assim, ser refletido, absorvido ou transmitido (ROSA, 2009; PONZONI; SHIMABUKURO, 2010). Os sensores utilizados no sensoriamento remoto

registram a reflectância do alvo, de grande importância na agricultura, pois diversas informações importantes a respeito do estado metabólico das culturas podem ser extraídas.

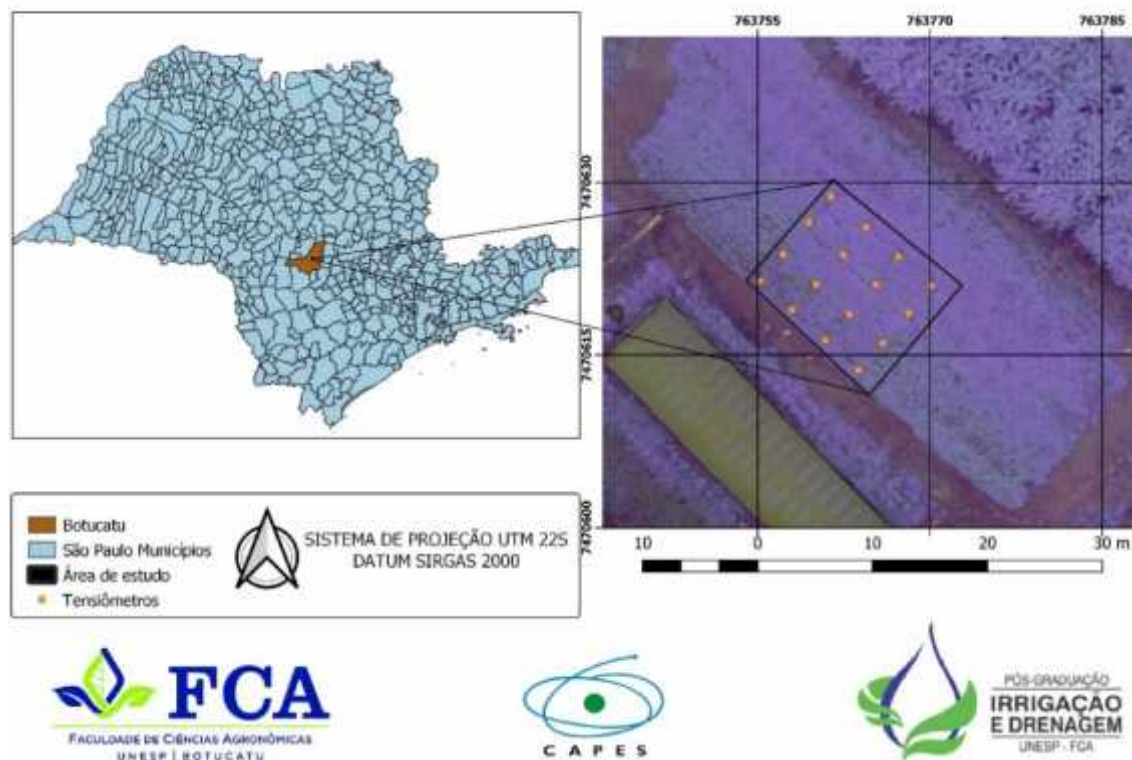
Recentemente, métodos estatísticos não lineares, como o aprendizado de máquinas (*machine learning*), vêm sendo aplicados a dados de sensoriamento remoto para a estimativa de parâmetros físicos. Algoritmos como *Random Forest* (BREIMAN, 2001), *Support Vector Machines* (KOVA EVIC; BAJAT; GAJIC, 2010; PRIORI; BIANCONI; CONSTANTINI, 2014), Redes Neurais Artificiais (AITKENHEAD et al., 2013; SILVEIRA et al., 2013) e vizinho próximo – k (MANSUY et al., 2014) têm sido bastante utilizados na agricultura de precisão.

O objetivo desta pesquisa foi determinar a umidade do solo a partir de imagens obtidas por meio de veículo aéreo não tripulado (VANT), utilizando algoritmos de *machine learning*, como *Support Vector Regression*, *Random Forests* e Redes Neurais Artificiais, tendo como referência, a medida de tensão de água no solo.

4 MATERIAL E MÉTODOS

4.1 Área de estudo

O estudo foi conduzido em uma parcela de 174 m², localizada na área experimental do Departamento de Engenharia Rural e Socioeconomia, da Faculdade de Ciências Agrônomicas (FCA) da Universidade Estadual Paulista “Júlio de Mesquita Filho” (Unesp) de Botucatu, SP, como pode ser observado na Figura 1.

Figura 1. Localização da área de estudo.

O solo da área é caracterizado como Nitossolo vermelho de textura argilosa, segundo o Mapa Pedológico do Estado de São Paulo (ROSSI, 2017). De acordo com a classificação de Köppen e Geiger (1928), o clima da região é subtropical úmido, Cfa, com duas estações bem definidas, verão quente e úmido e inverno seco. O relevo é plano com declividade de 1%. Na parcela, foi instalada a cultura do feijão (*Phaseolous vulgaris* L.), variedade Dama, com ciclo de 91 dias. A irrigação foi realizada via aspersão convencional com sobreposição dos aspersores de 12 x 12 m.

Na Figura 2, é possível visualizar os equipamentos utilizados para o aerolevanteamento. O VANT utilizado foi um Phantom 3® Professional (fabricante SZ DJI Technology Co., Shenzhen, Guangdong, China). O voo foi realizado de maneira autônoma a uma altitude de 120 m, com 60% de sobreposição lateral e 70% de

sobreposição frontal. O sensor utilizado foi uma MAPIR Survey 3W (fabricante MAPIR, Peau Productions, Inc., CA, USA). A MAPIR é uma câmera modificada com um filtro para registro no Infravermelho próximo (850 nm) e registro no Vermelho (660 nm) e Verde (550 nm). A câmera possui resolução de 12 megapixels (4032 x 3024) e produz imagens no formato JPG (8-bits) e RAW (12-bits). Foram utilizadas as configurações padrões de abertura, balanço de branco, Sensibilidade ISO (International Standards Organization), abertura do obturador e exposição, conforme recomendado pelo fabricante. Atualmente, esse sensor é o mais barato no mercado (\cong R\$ 6.000) quando comparado a outras câmeras multiespectrais disponíveis como a Parrot Sequoia (\cong R\$ 30.000) e Micasense RedEdge-M (\cong R\$ 50.000).

Figura 2. VANT e painel de calibração.



Um painel de calibração radiométrica foi utilizado para conversão dos números digitais em reflectância. Foram coletadas imagens antes do voo para calibração, e o processamento das imagens captadas foi realizado através do *software* Mapir Camera Control v. 10/16/2019. Para a calibração geométrica, foi utilizado um receptor GNSS RTK Kronos 200 (fabricante Horizon, Survey Instruments Ltd., Singapore) tendo sido coletados cinco pontos de controle distribuídos na área. Para a estimativa de umidade do solo, foram realizados dois levantamentos, o primeiro no dia 18 e o segundo no dia 20, ambos no mês de janeiro de 2020.

Os instrumentos para o monitoramento da tensão de água no solo

são apresentados na Figura 3. Utilizou-se uma bateria com 16 tensiômetros instalados na profundidade de 20 cm. A água nos tensiômetros foi mantida a um nível constante após as medições. A Realizou-se a leitura de tensão por meio de um tensímetro digital de agulha (fabricante Hidrodinâmica Tensiômetros, Piracicaba, Brasil). Retirou-se amostras de solo para a obtenção e cálculo da curva de retenção, utilizando a metodologia da câmara de pressão. Posteriormente, os valores de tensão foram ajustados por meio do modelo de Van Genuchten (1980) para obtenção dos valores do conteúdo volumétrico de água no solo (cm^3/cm^3). Os dados então foram multiplicados por 100 para a obtenção da umidade do solo em porcentagem.

Figura 3. Tensiómetro e tensímetro digital de agulha.

4.2 Support Vector Regression

Os algoritmos de *Support Vector Machines* também podem ser aplicados a problemas de regressão por meio da introdução de uma função de perda alternativa (SMOLA, 1996). Enquanto os algoritmos de *Support Vector Regression* (SVR) são uma generalização do problema de classificação encontrados nos classificadores do *Support Vector Machines*. Nesses algorítmicos, os erros são fixados de maneira a descartar pontos para a regressão de acordo com os pontos dentro do intervalo de confiança. O critério de otimização penaliza pontos de dados nos quais os valores y diferem de $f(x)$ mais que o erro ϵ .

A função genérica de estimativa do SVR pode ser descrita conforme a Equação 1:

$$f(x) = (\omega * \Phi(x)) + b \quad (1)$$

Em que, Φ é uma transformação não linear para um espaço dimensional maior. O

objetivo é encontrar um valor de ω e b de modo que os valores de x possam ser determinados minimizando o risco de regressão (R_{reg}) representado pela Equação 2 (WU; HO; LEE, 2004).

$$R_r = C \sum_{i=0}^l \Gamma(f(x_i) - y_i) + \frac{1}{2} \|w\|^2 \quad (2)$$

Em que, Γ é uma função de custo, C é uma constante e o vetor w pode ser escrito em termos dos pontos de dados (Equação 3):

$$w = \sum_{i=1}^l (a_i - a_i^*) \Phi(x_i) \quad (3)$$

O produto da transformação Φ pode ser estimado usando a função $k(x_i, x)$ chamada de função Kernel. A Função de Base Radial (RBF) é o Kernel mais utilizado e pode ser definida conforme a Equação 4:

$$R = e^{-\gamma |x - x_i|} \quad (4)$$

A função de custo mais usada na literatura é a e-insensitivo. Ela determina um

tubo delimitado pelos vetores de suporte para corte. Ela é resolvida em função de um valor de erro aceitável com base nos dados (Equação 5).

$$\Gamma(f(\mathbf{x}) - \mathbf{y}) = |f(\mathbf{x}) - \mathbf{y}| \geq \varepsilon \quad (5)$$

Mais detalhes sobre a teoria dos SVR podem ser encontrados em Smola e Scholkopf (2004) e Vapnik (1998).

4.3 Random Forests

O algoritmo *Random Forests* ou Florestas Aleatórias é um algoritmo de aprendizado supervisionado que utiliza o método de aprendizado em conjunto para classificação e regressão. No algoritmo, são criadas diversas árvores de decisão de maneira aleatória, formando algo

semelhante a uma floresta, onde cada árvore é utilizada para definição do resultado final. Tem-se uma coleção de árvores preditoras $h(x; k)$, $k=1, \dots, K$, onde x representa a vetor de entrada observado de comprimento p com um vetor aleatório \mathbf{X} e k são vetores aleatórios distribuídos de maneira idêntica (SEGAL, 2004).

O erro médio quadrático de generalização para qualquer preditor numérico $h(\mathbf{x})$ é descrito conforme a Equação 6 (BREIMAN, 2001):

$$E_{\mathbf{X}, \mathbf{Y}}(\mathbf{y} - h(\mathbf{X}))^2 \quad (6)$$

O preditor de *Random Forests* é formado, tomando a média sobre k das árvores $\{h(\mathbf{x}, k)\}$ conforme o número de árvores tende ao infinito (Equação 7):

$$E_{\mathbf{X}, \mathbf{Y}}(\mathbf{y} - \text{méd}_k h(\mathbf{X}, \theta_k))^2 \rightarrow E_{\mathbf{X}, \mathbf{Y}}(\mathbf{y} - E_{\theta} h(\mathbf{X}; \theta))^2 \quad (7)$$

A quantidade na direita é o erro de predição (generalização) para a *Random Forest* designado PE_f^* . O erro de predição de um árvore individual $h(\mathbf{X}; \theta)$ pode ser definido como (Equação 8):

$$PE_{\theta}^* = E_{\theta} E_{\mathbf{X}, \mathbf{Y}}(\mathbf{y} - h(\mathbf{X}; \theta))^2 \quad (8)$$

Supondo que para todos θ , a árvore não possui viés ($E_{\mathbf{Y}} = E_{\mathbf{X}} h(\mathbf{X}, \theta)$). Então, temos a Equação 9 (SEGAL, 2004):

$$PE_f^* \leq \bar{p} PE_{\theta}^* \quad (9)$$

Em que, \bar{p} é a correlação ponderada entre os resíduos $\mathbf{y} - h(\mathbf{X}; \theta)$ e $\mathbf{y} - h(\mathbf{X}; \theta')$ para θ, θ' que são vetores independentes distribuídos aleatoriamente.

As *Random Forests* funcionam construindo diversas árvores de decisão no processo de treinamento e dando como *output* o valor médio predito pelas árvores individuais. Cada árvore extrai uma amostra

aleatória do conjunto, de modo a adicionar mais um elemento de aleatoriedade evitando o *overfitting*. As árvores são rodadas em paralelo não havendo interação entre elas ao construir novas árvores, sendo classificadas como um algoritmo de *bagging* (ensacamento) e não um algoritmo de *boosting* (reforço) como as Redes Neurais.

4.4 Artificial Neural Network

Artificial Neural Networks, ou Redes Neurais Artificiais, são algoritmos que usam uma abordagem semelhante à estrutura do cérebro humano para tomada de decisão (MCCULLOCH; PITTS, 1943; ROSENBLATT, 1962; BISHOP, 1995). As variáveis de entrada independentes x_i ($i=1, \dots, d$) são transformadas em um conjunto de saída dependente y_1, \dots, y_i . O primeiro estágio da transformação é realizado pelo neurônio ou *Perceptron*. Os *inputs* são multiplicados por um parâmetro de peso w_i que simula os pesos sinápticos nas redes neurais

biológicas. Todas as entradas ponderadas serão somadas para dar uma entrada total (Equação 10):

$$u = \sum_{i=1}^d w_i x_i + b \quad (10)$$

Em que, um viés b é adicionado para fornecer um mecanismo para incluir outras influencias, normalmente definido como 1. O processo de aprendizagem começa com algum vetor de pesos arbitrários, sem perda de generalização, podemos assumir que esse seja como o vetor zero (BISHOP, 1995). Uma função de ativação vai definir o limiar aceitável para passar o valor da soma ponderada mais o valor do viés.

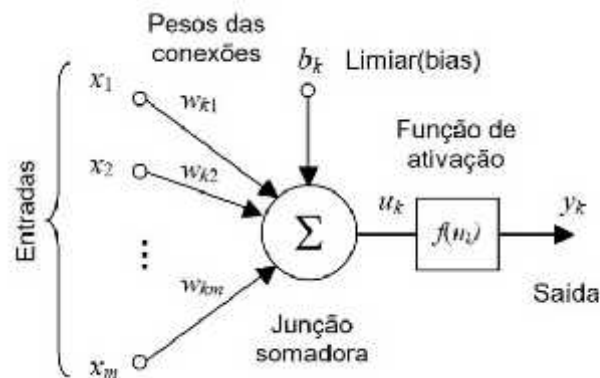
O resultado do neurônio é então enviado para uma função de ativação não

linear $f()$ que vai definir, com base no limiar, se esse vai ser ativado ou não (Equação 11).

$$z = f(u) \quad (11)$$

As redes normalmente usam a função sigmoide para o cálculo do limiar. Como esse tipo de rede opera através do neurônio, a partir dos *inputs*, para obter o *output*, são chamadas de *feed-foward* (BISHOP, 1995). As estruturas mais simples de rede possuem somente um neurônio como mostra a Figura 4, entretanto, as estruturas mais utilizadas são chamadas de *Multi Layer Perceptron* (MLP), uma vez que combinam diversas camadas de neurônios.

Figura 4. Modelo matemático de um neurônio artificial. Fonte: Adaptado de Haykin (2009).



A estrutura *Multi Layer Perceptron* é caracterizada por uma camada de entrada, camadas intermediárias chamadas de camadas ocultas e uma camada de saída. Assim que o processo de aprendizagem progride ao longo do MLP, os neurônios presentes na camada oculta começam a descobrir gradualmente as características salientes que caracterizam os dados de treinamento (HAYKIN, 2009).

O aprendizado da Rede é feito por meio do algoritmo de *Backpropagation*, que faz uso de uma técnica chamada gradiente descendente. Esse algoritmo ajusta os pesos com base na derivada destes em relação ao erro. A Rede é iniciada e os erros são

calculados no fim do processo, eles são então propagados de volta às camadas iniciais para ajuste dos pesos. O processo se repete diversas vezes até encontrar a combinação de pesos sinápticos que resultam no menor erro quando então o modelo converge.

Essa técnica mede o erro e a taxa de mudança do erro. Isso leva a grandes modificações quanto maior for o erro, conforme a inclinação diminui ao se aproximar de um mínimo, as mudanças nos pesos se tornam menores (PUJOL; PINTO, 2011). A média de todos os erros quadráticos (E) para a saída é computada para auxiliar na derivada. O descendente é baseado em um

gradiente no erro para todo o conjunto de dados segundo a Equação 12 (GROSSAN; ABRAHAM, 2011):

$$\Delta w_l(n) = -\eta * \frac{\delta}{z_w} + \alpha * \Delta w_l(n-1) \quad (12)$$

Em que, η é a taxa de aprendizado e α o momentum.

A utilização desse algoritmo já foi reportada em inúmeros estudos cujo objetivo principal foi a obtenção da estimativa de umidade do solo (JIANG; COTTON, 2004; AHMAD; KALRA; STEPHEN, 2010; KORNELSEN; COULIBALY, 2014; HASSAN-ESFAHANI et al., 2015).

4.5 Estratégia de modelagem

Foram utilizados *buffers* circulares ao redor da localização dos tensiômetros com raio de 0,5 m para extração dos valores médios de reflectância de cada banda por meio da ferramenta de estatística zonal do *software* QGIS 3.4. Esses dados foram utilizados como entrada nos algoritmos de *machine learning*. Para as variáveis independentes, utilizou-se as reflectâncias nas bandas do Verde, Vermelho, Infravermelho próximo e o Índice de Vegetação por Diferença Normalizada (NDVI). Empregou-se como A variável dependente, u a umidade do solo registrada por cada tensiômetro nos dois levantamentos realizados (18 e 20 de janeiro de 2020). A matriz de entrada tinha um formato de 32 x 4.

Os dados foram divididos em 70% para treinamento e 30% para teste das predições. Posteriormente, realizou-se uma normalização dos dados por meio do *Standard Scaler* (Equação 13), que padroniza as amostras, removendo a média e deixando a variância unitária com uma distribuição normal gaussiana. Esse processo é um requisito dos algoritmos de

machine learning para que os dados estejam na mesma escala.

$$z = \frac{(x-u)}{s} \quad (13)$$

Em que, x é uma amostra, u é a média das amostras de treinamento e s é o desvio padrão.

Os algoritmos utilizados nesse estudo foram implementados utilizando a linguagem Python 3.6 por meio da interface de usuário *Jupyter Notebook*, juntamente com as bibliotecas Pandas, Numpy e Scikit-Learn. Em relação aos hiperparâmetros selecionados para os estimadores, no algoritmo de SVR foi utilizado a função Kernel *Radial Basis Function*.

Para o *Random Forests* (RF), utilizou-se 100 árvores de decisão e o critério para otimização foi o '*mean Square Error*'. Já para as *Artificial Neural Networks* (ANN) foi utilizada uma arquitetura de MLP, inicializada com duas camadas ocultas com 100 neurônios cada com a função de ativação '*relu*'. A função de custo utilizada foi a "*Square Error*" e a taxa de aprendizado definida como 0,001, o otimizador selecionado foi o '*adam*', com a camada de saída sem função de ativação, gerado assim, um número contínuo.

4.6 Métricas estatísticas

Para a avaliação do desempenho dos algoritmos de *machine learning* empregados na determinação da estimativa de umidade do solo foram utilizadas as seguintes métricas estatísticas: Erro médio quadrático (*Root Mean Square Error* - RMSE) (Equação 14), Erro médio absoluto (*Mean Absolute Error* - MAE) (Equação 15), Erro

médio absoluto percentual (*Mean Absolute Percentage Error* - MAPE) (Equação 16) e Coeficiente de determinação (R^2) (Equação 17).

$$R = \sqrt{\frac{1}{n} \sum_{i=1}^n (y_i - x_i)^2} \quad (14)$$

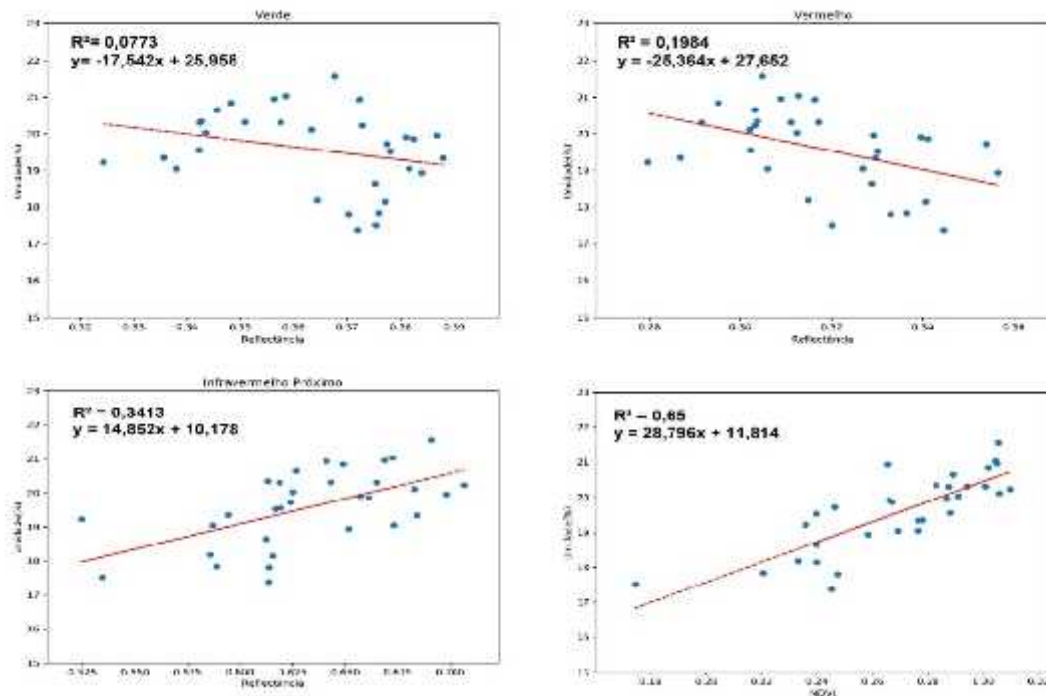
$$M = \frac{1}{n} \sum_{i=1}^n |y_i - x_i| \quad (15)$$

$$M = \frac{1}{n} \sum_{i=1}^n \left| \frac{y_i - x_i}{x_i} \right| * 100 \quad (16)$$

$$R^2 = \frac{\sum_{i=1}^n (y_i - x_i)^2}{\sum_{i=1}^n (y_i - \bar{y})^2} \quad (17)$$

Em que, y_i são os valores observados de umidade do solo, x_i são os valores simulados, n é o número de observações e \bar{y} é a média das observações.

Figura 5. Gráficos de dispersão demonstrando a correlação entre a umidade do solo e as bandas espectrais.



Também é possível identificar uma baixa correlação positiva entre o Infravermelho próximo e a umidade do solo.

5 RESULTADOS E DISCUSSÃO

5.1 Umidade do solo e reflectância

Na Figura 5, é possível observar as correlações entre as bandas espectrais (Verde, Vermelho e Infravermelho Próximo), o NDVI e a umidade do solo estimada indiretamente pelos tensiômetros. Observa-se uma correlação negativa entre a banda do vermelho e a umidade do solo. Quando a planta não está estressada, a radiação na faixa do vermelho é refletida com menor intensidade devido à absorção da clorofila pelos carotenos e xantofilas, ou seja, áreas com elevado teor de umidade no solo têm menor refletância nessa faixa do espectro eletromagnético.

Isso ocorre porque onde há maior disponibilidade de água, a planta tem um maior desenvolvimento vegetativo,

umentando a refletância no Infravermelho próximo devido ao maior Índice de Área Foliar (IAF). A banda do verde não apresentou correlação significativa com a umidade do solo na profundidade analisada. Esses resultados são similares aos encontrados por Aboutaleb et al. (2019), que observaram a mesma tendência de correlação entre a umidade do solo e essa banda espectral em camadas de solo a partir de 45 cm de profundidade.

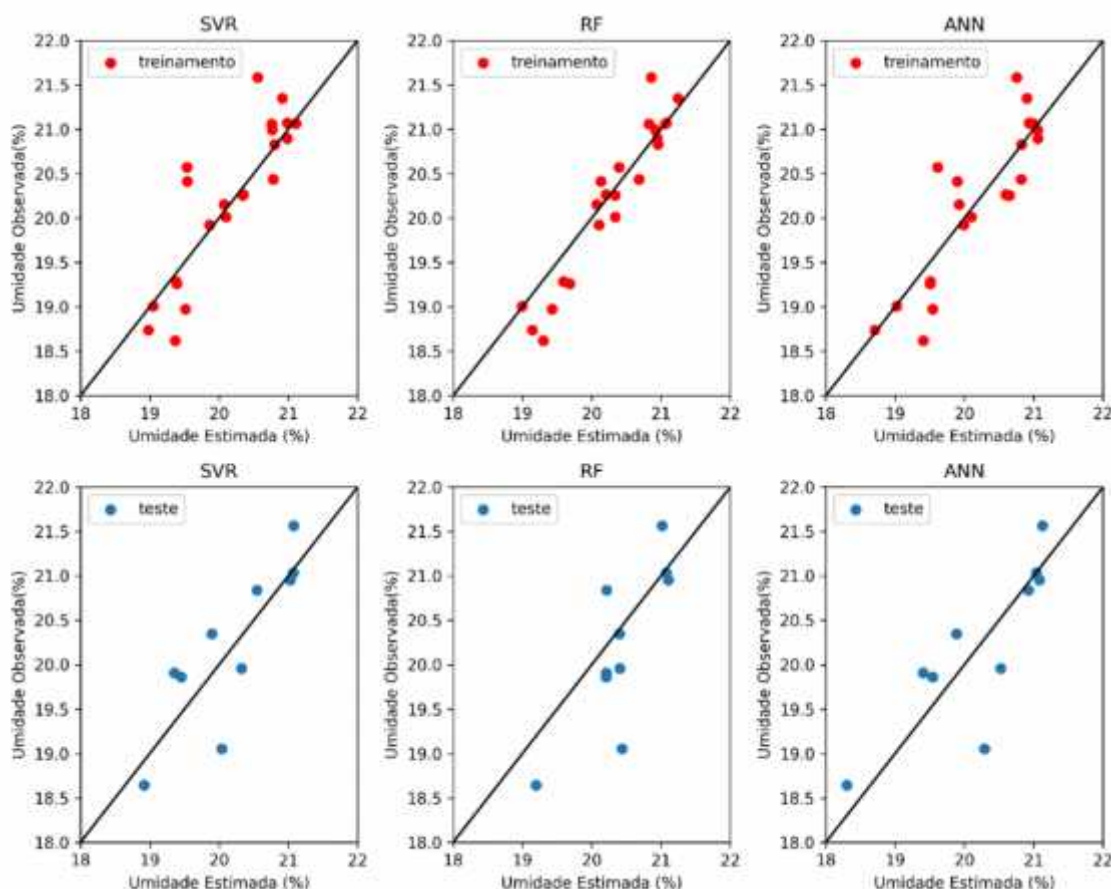
É possível notar que existe uma alta correlação positiva entre o NDVI e a umidade do solo (Figura 5). O NDVI está relacionado com a quantidade de umidade no solo, pois uma planta sem restrições hídricas pode alcançar pleno vigor vegetativo, apresentando altos valores de NDVI, enquanto áreas com baixo NDVI,

geralmente associadas a maior presença de solo exposto nos pixels, podem indicar que as plantas estão sofrendo estresse hídrico devido à baixa disponibilidade hídrica no solo (menor umidade). A câmera MAPIR tende a apresentar valores baixos de NDVI como reportado por Gomes et al. (2021).

5.2 Performance dos estimadores de machine learning

A Figura 6 apresenta a umidade estimada pelos algoritmos de *machine learning* e a umidade observada nos tensiômetros para os conjuntos de treinamento e teste. Todos os algoritmos tiveram desempenho satisfatório com RMSE < 1 vol. % na estimativa de umidade do solo.

Figura 6. Gráficos de dispersão dos valores de tensão de água no solo observados pelos tensiômetros versus valores estimados pelos algoritmos de *machine learning*.



A Tabela 2 apresenta um resumo das estatísticas avaliadas utilizando os dados de teste. Como pode ser observado, o algoritmo SVR foi melhor em relação ao RF e ANN. Isso pode ser confirmado pela distribuição dos pontos em relação a linha 1:1, onde o agrupamento foi maior no algoritmo SVR. O SVR apresentou um RMSE de 0,46 vol. % e

o valor do MAPE foi de 4,59 %. O ajuste da regressão foi de 0,71 e o MAE de 0,39 vol. %. O desempenho do algoritmo de ANN foi similar ao SVR com um RMSE de 0,54 vol. % e MAPE de 4,23 %, o R^2 foi de 0,60 e o MAE de 0,40 vol. %. Já o RF apresentou um RMSE de 0,57 vol. % e MAPE de 4,23 %. O R^2 foi de 0,55 com MAE de 0,42 vol. %.

Tabela 1. Resumo das estatísticas dos valores observados versus valores estimados em porcentagem de umidade do solo.

	RMSE ¹ (vol %)	MAE (vol %)	MAPE (%)	R ²
SVR ²	0,46	0,39	4,59	0,71
RF	0,57	0,42	4,23	0,55
ANN	0,54	0,40	4,75	0,60

De maneira geral, os modelos conseguiram estimar de maneira precisa o valor da umidade do solo utilizando as bandas do verde, vermelho, infravermelho próximo e o NDVI. Valores muito baixos de erro normalmente estão associados ao fenômeno de *overfitting*, porém ao avaliar os algoritmos nos dados de validação (teste), o R^2 baixo demonstra que o ajuste não foi perfeito, característica essencial de *overfitting* em no qual o R^2 apresenta valores iguais ou muito próximos a 1. Entretanto, os valores de umidade não apresentaram amplitude significativa nos levantamentos (1,3%) o que pode explicar o valor baixo de RMSE.

Os resultados apresentados são semelhantes aos encontrados por Araya et al. (2021), que obtiveram valores para MAE de 3,77 vol. % ao aplicar os mesmos algoritmos empregados neste trabalho em dados multiespectrais coletados pelo sensor Parrot Sequoia (fabricante Parrot SA, Paris, France) em seis levantamentos com 406 amostras de umidade do solo. Ge et al. (2019) também encontraram resultados semelhantes (RMSE = 1,47 vol. %) ao utilizar os algoritmos RF e *Extreme*

Learning Machine em 70 amostras de umidade do solo coletadas em um levantamento com uma câmera hiperespectral. Isso demonstra o potencial da utilização de métodos de *machine learning* em dados coletados via sensoriamento remoto e sua aplicação no monitoramento do conteúdo de umidade no solo o que, conseqüentemente, permite o manejo adequado da irrigação, apresentando-se como uma opção viável à prática da agricultura de precisão.

5.3 Mapas de umidade do solo

A Figura 7 apresenta os mapas gerados pelos algoritmos de *machine learning* nos levantamentos realizados nos dias 18 e 20 de janeiro. Todos os mapas apresentam valores menores de umidade no canto inferior esquerdo, o que está de acordo com os dados registrados nos tensiômetros. É possível identificar que o algoritmo RF suavizou os dados, gerando um mapa mais uniforme. Já o algoritmo SVR gerou um mapa de umidade do solo com grande variação espacial para ambos levantamentos. O algoritmo ANN teve o pior desempenho

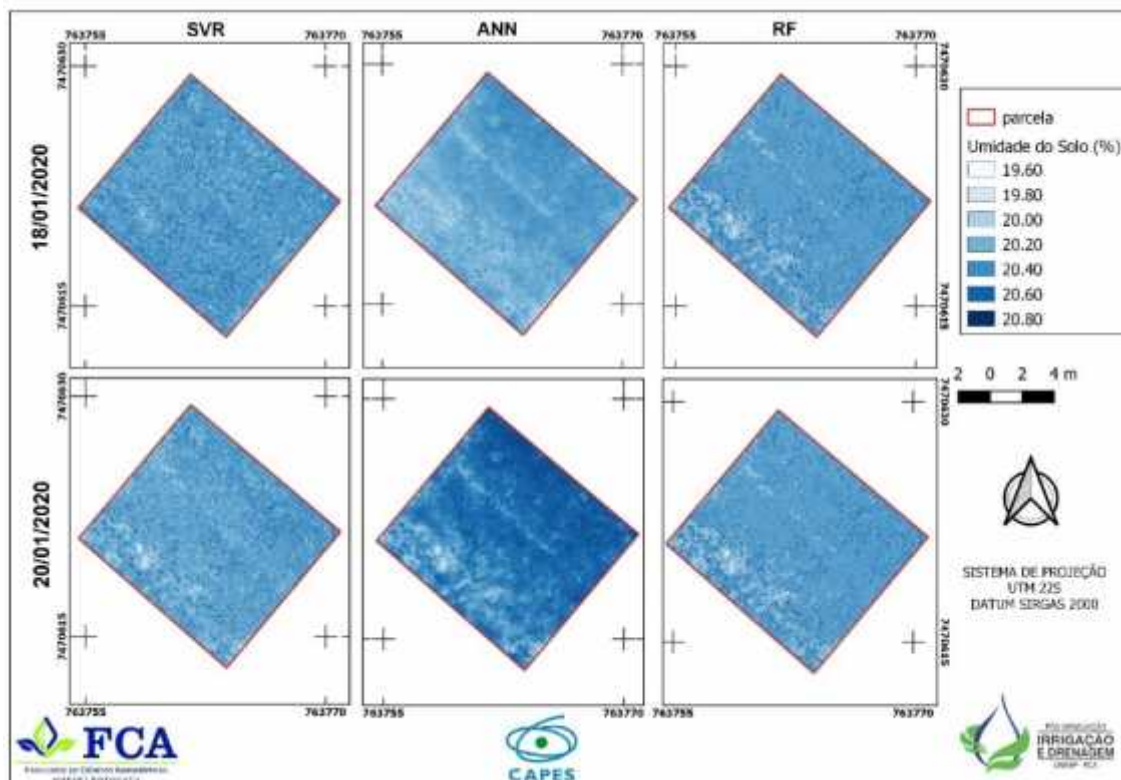
¹ Root Mean Square Error (RMSE); Mean Absolute Error (MAE); Mean Absolute Percentage Error (MAPE), Coeficiente de determinação (R^2).

² Support Vector Regression (SVR); Random Forests (RF) e Artificial Neural Network (ANN).

na estimativa de umidade do solo, apresentando uma maior generalização dos

valores, sendo inclusive, possível identificar *clusters* com posições mais definidas.

Figura 7. Mapas de umidade do solo gerados pelos algoritmos de *machine learning* nos dois levantamentos.



Os mapas foram gerados em alta resolução com pixel de 12 cm, o que pode não ser tecnicamente viável para manejo o da irrigação, porém os dados podem ser reamostrados por meio da técnica de *downsampling* que permite a redução da resolução espacial. Outra alternativa é a realização de voos em altitude superior a 120 m, que é mais interessante para levantamentos em grandes áreas como as ocupadas por sistemas de irrigação por pivô central. Entretanto, ressalta-se que o objetivo deste trabalho foi demonstrar a precisão dos algoritmos de *machine learning* na estimativa de umidade do solo, utilizando para tal, dados coletados por um sensor óptico de baixo custo. A possibilidade de coletar dados em qualquer intervalo de tempo utilizando um VANT, somada à capacidade de reconhecimento de padrões

por algoritmos de *machine learning*, configura-se uma excelente ferramenta para a aquisição remot de dados de umidade do solo.

6 CONCLUSÕES

Foi possível demonstrar nesse estudo, a utilidade dos algoritmos de *machine learning* para a estimativa da umidade do solo a partir de imagens multiespectrais de alta resolução coletadas por VANT. As bandas do Vermelho e Infravermelho próximo foram as mais correlacionadas com a tensão de água no solo na profundidade de 20 cm. A performance dos algoritmos SVR e ANN foram similares, apresentando pouca diferença em relação ao ajuste da regressão.

O melhor algoritmo para a estimativa de umidade do solo, com base nos dados analisados, foi o SVR, que apresentou um RMSE de 0,46 vol. % na estimativa de umidade do solo nos dados de teste e R^2 de 0,71. Os mapas gerados pelos algoritmos demonstram a alta variabilidade espacial da umidade do solo e podem ser utilizados para o monitoramento da mesma em qualquer escala temporal, sendo a técnica aqui apresentada limitada somente por condições climatológicas que podem impedir o levantamento de dados por meio do uso de um VANT.

A performance dos algoritmos de *machine learning*, em especial as Redes Neurais, é influenciada pelo tamanho dos dados de treinamento, podendo haver *overfitting* em conjuntos de dados não tão extensos, o que representa uma limitação dos resultados aqui apresentados. Todavia,

destaca-se que novas campanhas podem ser realizadas para a coleta de dados, e esses dados utilizados para retreinar os modelos, aumentando assim, cada vez mais a precisão dos mesmos.

Esse estudo fornece evidências de que o *machine learning* pode ser empregado na estimativa, por método indireto, do teor de água do solo, sendo possível obter resultados de alta acurácia mesmo utilizando câmeras de baixo custo acopladas ao veículo aéreo não tripulado.

7 AGRADECIMENTOS

Os autores agradecem à CAPES pelo auxílio DS 88882.433001/2019-01 e ao CNPQ pelo auxílio 131325/2020-5, essenciais para a realização deste trabalho.

8 REFERÊNCIAS

ABOUTALEBI, M.; ALLEN, L. N.; TORRES-RUA, A. F.; MCKEE, M.; COOPMANS, C. Estimation of soil moisture at different soil levels using machine learning techniques and unmanned aerial vehicle (UAV) multispectral imagery". *In: SPIE DEFENSE + COMMERCIAL SENSING*, 6., 2019, Baltimore. **Proceedings** [...]. Baltimore: Autonomous Air and Ground Sensing Systems for Agricultural Optimization and Phenotyping, 2019. p. 1-12.

AHMAD, S.; KALRA, A.; STEPHEN, H. Estimating soil moisture using remote sensing data: A machine learning approach. **Advances in Water Resources**, Iowa City, v. 33, p. 69-80, 2010.

AITKENHEAD, M. J.; COULL, M.; TOWERS, W.; HUDSON, G.; BLACK, H. I. J. Prediction of soil characteristics and colour using data from the National Soils Inventory of Scotland. **Geoderma**, Amsterdã, v. 200/201, p. 99-107, 2013.

ARAYA, S. N.; FRYJOFF-HUNG, A.; ANDERSON, A.; VIERS, J. H.; GHEZZEHEI, T. A. Advances in soil moisture retrieval from multispectral remote sensing using unoccupied aircraft systems and machine learning techniques. **Hydrologic Earth Systems Science**, Gottingen, v. 25, p. 2739-2758, 2021.

- ARRUDA, L. E. V.; FIGUEIRÊDO, V. B.; LEVIEN, S. L. A.; MEDEIROS, J. F. Desenvolvimento de um tensiômetro digital com sistema de aquisição e armazenamento de dados. **Irriga**, Botucatu, v. 1, n. 1, p. 11-20, 2017.
- BISHOP, C. M. **Neural Networks for Pattern Recognition**. Oxford: Oxford University Press, 1995.
- BREIMAN, L. Random forests. **Machine Learning**, New Jersey, v. 45, p. 5-32, 2001.
- BRITO, A. S.; LIBARDI, P. L.; MOTA, J. C. A.; MORAES, S. O. Desempenho do tensiômetro com diferentes sistemas de leitura. **Revista Brasileira de Ciência do Solo**, Viçosa, v. 33, n. 1, p. 17-24, 2009.
- BRITO, A. S.; LIBARDI, P. L.; MOTA, J. C. A.; KLEIN, V. A. Variação diurno-noturna do potencial mátrico e gradiente de potencial total da água no solo. **Revista Brasileira de Ciência do Solo**, Viçosa, v. 38, n. 1, p. 128-134, 2014.
- CAMARGO, A. P.; GROHMANN, F.; CAMARGO, M. B. P. Tensiômetro simples de leitura direta. **Pesquisa Agropecuária Brasileira**, Brasília, DF, v. 17, n. 12, p. 1763-72, 1982.
- DOBRIYAL, P.; QURESHI, A.; BADOLA, R.; HUSSAIN, S. A. A review of the methods available for estimating soil moisture and its implications for water resource management. **Journal of Hydrology**, Amsterdam, v. 458-459, p. 110-117, 2012.
- DUKES, M. D.; ZOTARELLI, L.; MORGAN, K. T. Use of irrigation technologies for vegetable crops in Florida. **HortTechnology**, Alexandria, v. 20, n. 1, p. 133-142, 2010.
- FREITAS, W. A.; CARVALHO, J. A.; BRAGA, R. A.; ANDRADE, M. J. B. Manejo da irrigação utilizando sensor da umidade do solo alternativo. **Revista Brasileira de Engenharia Agrícola e Ambiental**, Campina Grande, v. 16, n. 3, p. 268-274, 2012.
- GE, X.; WANG, J.; DING, J.; CAO, X.; ZHANG, Z.; LIU, J.; LI, X. Combining UAV-based hyperspectral imagery and machine learning algorithms for soil moisture content monitoring. **Environmental Science**, Amsterdã, v. 7, n. 6929, p. 1-27, 2019.
- GOMES, A. P. A.; QUEIROZ, D. M.; VALENTE, D. S. M.; PINTO, F. A. C.; ROSAS, J. T. F. Comparing a single-sensor câmera with a multisensor câmera for monitoring coffee crop using unmanned aerial vehicles. **Engenharia Agrícola**, Jaboticabal, v. 41, n. 1, p. 87-97, 2022.
- GROSSAN, C.; ABRAHAM, A. **Intelligent Systems: A modern approach**. Berlin: Springer, 2011.
- HASSAN-ESFAHANI, L.; TORRES-RUA, A.; JENSEN, A.; MCKEE, M. Assessment of surface soil moisture using high-resolution multi-spectral imagery and artificial neural networks. **Remote Sensing**, Basel, v. 7, n. 3, p. 2627-2646, 2015.
- HAYKIN, S. **Neural Networks: A Comprehensive Foundation**. 2. ed. New Jersey: Prentice Hall, 1999.

HUISMAN, J. A.; HUBBARD, S. S.; REDMAN, J. D.; ANNAN, A. P. Measuring soil water content with ground penetrating radar: a review. **Vadose Zone Journal**, Madison, v. 2, n. 4, p. 476-491, 2003.

JENSEN, J. R. **Remote sensing of the environment: an earth resource perspective**. 2. ed. New Delhi: Pearson Education, 2009.

JIANG, H.; COTTON, W. Soil moisture estimation using an artificial neural network: A feasibility study. **Canadian Journal of Remote Sensing**. Quebec, v. 30, n. 5, p. 827-839, 2004.

KÖPPEN, W.; GEIGER, R. **Klimate der Erde**. Gota: Justus Perthes, 1928.

KOVA EVIC, M.; BAJAT, B.; GAJI, B. Soil type classification and estimation of soil properties using support vector machines. **Geoderma**, Amsterdã v. 154, n. 3/4, p. 340-347, 2010.

KORNELSEN, K. C.; COULIBALY, P. Root-zone soil moisture estimation using data-driven methods. **Water Resources Research**, New Jersey, v. 50, p. 2946-2962, 2014.

LIBARDI, P. L. **Dinâmica da água no solo**. São Paulo: EDUSP, 2005.

MANSUY, N.; THIFFAULT, E.; PARÉ, D.; BERNIER, P.; GUINDON, L.; VILLEMAIRE, P.; POIRIER, V.; BEAUDOIN, A. Digital mapping of soil properties in Canadian managed forests at 250m of resolution using the k-nearest neighbor method. **Geoderma**, Amsterdã, v. 235/236, p. 59-73, 2014.

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

MONTESANO, F. F.; SERIO, F.; MININNI, C.; SIGNORE, A.; PARENTE, A.; SANTAMARIA, P. Tensiometer-based irrigation management of subirrigated soilless tomato: effects of substrate matric potential control on crop performance. **Frontiers in Plant Science**, Lausanne, v. 6, n. 1150, p. 1-11, 2015.

NOVO, E. M. L. **Sensoriamento Remoto: princípios e aplicações**. Sao Paulo: Edgard Blucher, 2008.

OR, D. Who invented the tensiometer? **Soil Science Society of America Journal**, New Jersey, v. 65, n. 1, p. 1-3, 2001.

PONZONI, F. J.; SHIMABUKURO, Y. E. **Sensoriamento Remoto no estudo da vegetação**. São José dos Campos: Parêntese, 2010.

PRIORI, S.; BIANCONI, N.; CONSTANTINI, E. A. C. Can γ -radiometrics predict soil textural data and stoniness in different parent materials? A comparison of two machine learning methods. **Geoderma**, Amsterdã, v. 226/227, p. 354-364, 2014.

PUJOL, J. C. F.; PINTO, J. M. A. A Neural-network approach to fatigue life prediction. **International Journal of Fatigue**, Amsterdã, v. 33, n. 3, p. 313-322, 2011.

ROSA, R. **Introdução ao sensoriamento remoto**. 7. ed. Uberlândia: EDUFU, 2009.

ROSENBLATT, F. **Principles of neurodynamics**: perceptrons and the theory of brain mechanisms. New York: Spartan Books, 1962.

ROSSI, M. **Mapa pedológico do estado de São Paulo**: revisado e ampliado. São Paulo: Instituto Florestal, 2017.

SEGAL, M. R. **Machine Learning Benchmarks and Random Forest Regression**. San Francisco: University of California, 2004. Disponível em: <https://escholarship.org/uc/item/35x3v9t4>. Acesso em: 09 fev. 2022.

SILVEIRA, C. T.; OKA-FIORI, C.; SANTOS, L. J. C.; SIRTOLI, A. E.; SILVA, C. R.; BOTELHO, M. F. Soil prediction using artificial neural networks and topographic attributes. **Geoderma**, Amsterdã, v. 195/196, p. 165-172, 2013.

SMOLA, J. **Regression estimation with support vector learning machines**. Dissertação (Mestrado em Física) – Technische Universität München, Munique, 1996.

SMOLA, A. J.; SCHOLKOPF, B. **A tutorial on support vector regression**. London: Royal Holloway College, 2004.

TEIXEIRA, A. S.; COELHO, S. L. Desenvolvimento e calibração de um tensiômetro eletrônico de leitura automática. **Engenharia Agrícola**, Jaboticabal, v. 25, n. 2, p. 367-376, 2005.

THALHEIMER, M. A low cost electronic tensiometer system for continuous monitoring of soil water potential. **Journal of Agricultural Engineering**, Pavia, v. 44, n. 3, p. 114-119, 2013.

VAN GENUCHTEN, M. T. A closed form equation for predicting the hydraulic conductivity of unsaturated soils. **Soil Science Society of America Journal**, Madison, v. 44, n. 5, p. 892-898, 1980.

VAPNIK, V. **Statistical Learning Theory**. New York: Springer, 1998.

WU, C. H.; DHO, J. M.; LEE, D. T. Travel-time prediction with support vector regression. **IEEE Transactions on intelligent transportation systems**, Blacksburg, v. 5, n. 4, p. 276-281, 2004.