

# Air temperature estimation techniques in Minas Gerais state, Brazil, Cwa and Cwb climate regions according to the Köppen-Geiger climate classification system

## Técnicas de estimativa da temperatura do ar no estado de Minas Gerais, Brasil, em regiões de clima Cwa E Cwb segundo sistema de classificação climática de Köppen-Geiger

Pietros André Balbino dos Santos<sup>1\*</sup>, Cassio Augusto Ussi Monti<sup>2</sup>, Luiz Gonsaga de Carvalho<sup>1</sup>,  
Wiliam Soares Lacerda<sup>3</sup>, Felipe Scherz<sup>1</sup>

<sup>1</sup>Universidade Federal de Lavras/UFLA, Departamento de Engenharia Agrícola, Lavras, MG, Brasil

<sup>2</sup>North Carolina State University, Department of Forestry and Environmental Resources, Raleigh, NC, U.S.A.

<sup>3</sup>Universidade Federal de Lavras/UFLA, Departamento de Automática, Lavras, MG, Brasil

\*Corresponding author: [pietros.balbino@gmail.com](mailto:pietros.balbino@gmail.com)

Received in September 1, 2020 and approved in February 10, 2021

### ABSTRACT

Air temperature significantly affects the processes involving agricultural and human activities. The knowledge of the temperature of a given location is essential for agricultural planning. It also helps to make decisions regarding human activities. However, it is not always possible to determine this variable. It is necessary to make a precise estimate, using methods that are capable of detecting the existing variations. The aim of this study was to develop models of multiple linear regression (MLR), artificial neural network (ANN), and random forest (RF) to estimate the mean (Tmean), maximum (Tmax), and minimum (Tmin) monthly air temperatures as a function of geographic coordinates and altitude for different localities in Minas Gerais state, Brazil, with climatic classification Cwa or Cwb. The average monthly data (Tmean, Tmax, and Tmin), over a period of 30 years, were collected from 20 climatological stations. The MLR was able to estimate the Tmax with accuracy. However, the predictive capacity of estimating Tmean and Tmin was low. The algorithms RF and ANN were used to estimate Tmean, Tmax, and Tmin with high accuracy. The best results were obtained using the RF model.

**Index terms:** Artificial neural network; random forest; multiple linear regression; geographic coordinates.

### RESUMO

A temperatura do ar afeta significativamente os processos que envolvem atividades agrícolas e humanas. O conhecimento da temperatura de um determinado local é fundamental para o planejamento agrícola. Também ajuda a tomar decisões sobre as atividades humanas. No entanto, nem sempre é possível determinar essa variável. É necessário fazer uma estimativa precisa, utilizando métodos que sejam capazes de detectar as variações existentes. O objetivo deste estudo foi desenvolver modelos de regressão linear múltipla (RLM), rede neural artificial (RNA) e floresta aleatória (FA) para estimar a temperatura média (Tmean), máximo (Tmax), e mínimo (Tmin) mensal do ar em função de coordenadas geográficas e altitude para diferentes áreas do Estado de Minas Gerais, Brasil, com classificação climática Cwa ou Cwb. Os dados médios mensais (Tmean, Tmax e Tmin), ao longo de um período de 30 anos, foram coletados em 20 estações climatológicas. O RLM foi capaz de estimar o Tmax com precisão. Porém, a capacidade preditiva de estimar Tmean e Tmin foi baixa. Os algoritmos FA e RNA foram usados para estimar Tmean, Tmax e Tmin com alta precisão. Os melhores resultados foram obtidos com o modelo RF.

**Termos para indexação:** Rede neural artificial; floresta aleatória; regressão linear múltipla; coordenadas geográficas.

## INTRODUCTION

It is important to monitor the meteorological elements to achieve proper growth and yield of crops. Efficient monitoring can help in evapotranspiration estimates, irrigation planning, pest and disease risk zoning, animal comfort index mapping, etc. One of

the most important meteorological elements is air temperature, which influences plant physiology. Changes in air temperature can lead to change in the growth and development of plants (Benlloch-González et al., 2016; Cardoso et al., 2012; Wahid et al., 2007). The air temperature influences various physiological processes occurring in a plant, such as the speed of chemical reactions

(Benavides et al., 2007) that occur in the temperature range of 0–40 °C. The extent of influence exerted depends on the plant species. When the air temperature exceeds the ideal range for each species, morphological, physiological, and biochemical changes may be induced, leading to adverse effects on plant growth (Wahid et al., 2007). Studies on the characterization of air temperature, precipitation, and the climatic classification of the regions where agriculture predominates should be conducted to improve crop yields (Cardoso et al., 2015; Costa et al., 2012).

In coffee crop science, one of the main crop types grown in the Minas Gerais State, Brazil (Compahia Brasileira de Abastecimento - CONAB, 2020), the optimum mean annual temperature falls in the range of 18 – 23 °C for the proper growth of *C. Arabica* specie. The optimum temperature falls in the range of 22 – 26 °C for the proper growth of *C. Canephora* (Damatta et al., 2018). Temperatures that fall outside this range influence the growth and yields of the crops. When the temperature is extremely low, the activity of the coffee crop reduces, and the photosynthetic performance is noticeably affected. The net photosynthetic activity ceases almost completely (Batista-Santos et al., 2011; Partelli et al., 2009). On the other hand, very high temperatures may cause a decrease in the net photosynthetic rates of the leaves (Cannell, 1985). The ideal temperature interval produces a high crop yield over the years. The temperature outside the optimal range results in reduced crop yield. Therefore, it is important to determine the mean air temperature and the extreme temperatures (maximum and minimum). Furthermore, considering the characteristics of the relief and location of the Minas Gerais State, the accurate estimation of extreme temperatures is important because the state exhibits topographic conditions that allow the formation of frosts on an annual basis in the southern region. The maximum temperatures (40 – 42 °C) are recorded in the northern regions of the state.

The mean, maximum, and minimum air temperatures can be monitored on a daily basis in weather stations. However, in the Minas Gerais region, the coverage of the official network of surface weather stations is limited. Besides, interruptions and errors in the database generated by these stations are quite common. The errors can be attributed to reading errors, damaged devices, and other unintended observational problems (Dumedah; Coulibaly, 2011; Mwale; Adeloje; Rustum, 2012). These factors limit climatic studies, e.g., studies on the climatic characterization of the region and studies on meteorological elements that slow down the development of agriculture.

Considering the fact that the average monthly air temperature varies with geographic coordinates and

altitude, several researchers working in different regions of Brazil have been trying to develop techniques and models for estimating the air temperature. The multiple linear regression (MLR) model considers the latitude, longitude, and altitude of the location as independent variables (Alvares et al., 2013; Cargnelutti Filho; Maluf; Matzenauer, 2008; Pezzopane et al., 2004; Sedyama; Melo Júnior, 1998). These estimates have been made with different levels of precision and accuracy. However, the development of new tools such as the Artificial Neural Network and Random Forests technique can maximize the performance, precision, and accuracy of estimating the air temperature.

The new techniques have been developed with the aim of achieving higher accuracy during the estimation of variables. The Artificial Neural Network (ANN) is a promising and effective tool for non-linear modeling and complex time-series. It has been used in different fields of science such as medicine (Muhammad et al., 2019), hydrology (Asadi et al., 2019), and agriculture (De Oliveira Aparecido et al., 2020). The ANN model is a mathematical model in which the architecture is analogous to brain functioning. The interconnecting processing elements are arranged in several layers (Kumar; Raghuwanshi; Singh, 2011). The ANN method helps understand and generalize the relationships between complex datasets. This expands the scope of the application of the method (Wu; Dandy; Maier, 2014).

ANNs have been used for the estimation of meteorological variables with good accuracy. Estimation of reference evapotranspiration (Antonopoulos; Antonopoulos, 2017; Kumar; Raghuwanshi; Singh, 2011), solar radiation (Bou-Rabee et al., 2017), and air temperature (Moreira; Cecílio, 2016) have been carried out using this technique. It is important to conduct this study to verify the applicability of the ANN method for estimating the mean, maximum, and minimum air temperature. The efficiency of the technique has been investigated. Reports on the use of ANNs (used to estimate the temperature in the region under study) are scarce.

The Random Forest (RF) is non-parametric statistical data modeling methods (Breiman, 2001). The models have been used to analyze data in different fields of science, such as medicine (Xie et al., 2020), biology (Fabris et al., 2018), and geoprocessing (Vogels et al., 2017). According to James et al. (2013), decision trees detect non-linear relationships in the evaluated system when the use of linear relationships, e.g., linear regression analysis, is restricted. According to Seyedhosseini and Tasdizen (2015), RF is a classification and regression technique used to grow ensemble decision trees such that the correlation between the trees remains as low as

possible. This condition can be achieved by the method of bootstrap sampling. In this method, resamples are replaced by simulating a single random sample. It must represent samples taken from the original population. Data from previously conducted analytical experiments are required to enhance the predictive and generalization abilities (Hesterberg et al., 2002).

RF has also been adopted to predict meteorological variables such as solar radiation (Benali et al., 2019) and air temperature (Noi; Degener; Kappas, 2017). RF has been found to be a more efficient predicting tool compared to other tools like ANN (Benali et al., 2019; Zhou et al., 2016). The RF is still little applied, and the interest in this predictive tool is increasing as it exhibits a good practical performance (Scornet, 2016). Therefore, it is important to evaluate the RF potential for estimating air temperature and to compare it with different methods.

The objective of this study was to develop and compare the performances of multiple linear regression (MLR), Artificial Neural Networks (ANN), and Random Forests (RF) models for estimating the mean, maximum, and minimum monthly air temperatures using input variables such as geographical coordinates and altitude for different areas in the Minas Gerais State with climatic classification Cwa or Cwb (Köppen; Geiger, 1928).

## MATERIAL AND METHODS

### Study area and data sources

The present study was developed for municipalities in the Minas Gerais state that are within the regions classified as Cwa (humid temperate climate with dry winter and hot summer) and Cwb (humid temperate climate with dry winter and moderately hot summer). This classification was proposed by Köppen and Geiger (1928) (Figure 1). This Climatic Classification Systems (CMS) was developed by Köppen in 1918, and its most popular version was published in 1928 in collaboration with Rudolf Oskar Robert Williams Geiger. The Köppen and Geiger (1928) CMS a simple and comprehensive system, and hence it is widely used. The mean annual rainfall recorded in the region under study is 1379 mm (Brasil, 1992). The study was limited to the areas classified as Cwa and Cwb. The aim was to determine the maximum efficiency of the models tested. Highly accurate data were obtained when the models were used in regions exhibiting similar climatic characteristics.

According to De Sá Júnior et al. (2012), the regions classified as Cwa and Cwb represent 21% and 11% of the area of the Minas Gerais state, respectively. There are 20 climatological stations located in the region under study.

The regions fall under the realm of the national network of climatological stations (National Institute of Meteorology (INMET)). The respective geographical coordinates and climatic classification have been presented in Table 1. The average monthly data (mean (Tmean), maximum (Tmax), and minimum (Tmin) air temperature) over a period of 30 years, from 1987 to 2017, of each conventional station were used for the studies. The data were extracted from the Meteorological Database for Teaching and Research - BDMEP of INMET. Although some locations do not have a record of 30 years of data (Table 1), all stations presented more than 90% of the consistent data.

### Multiple linear regression (mlr) method

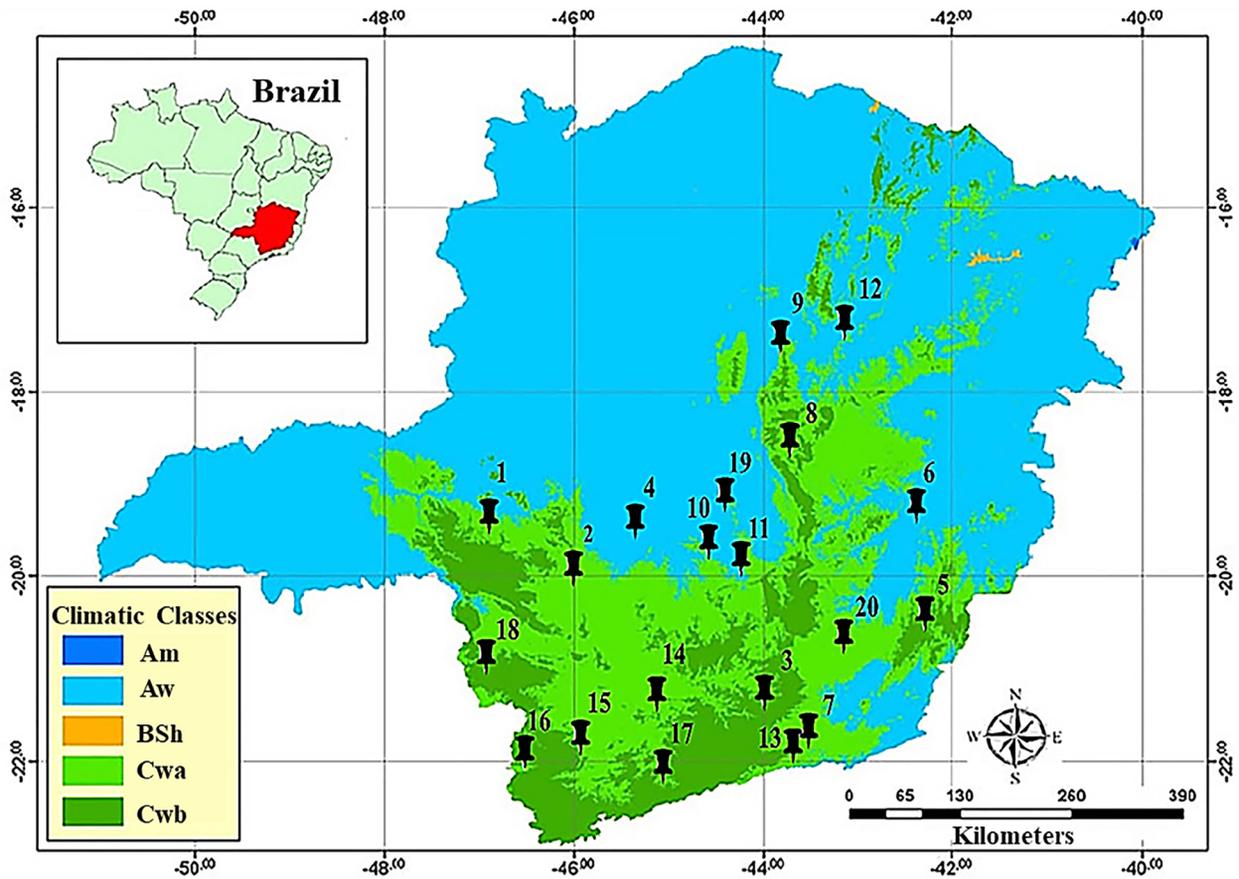
Based on the independent variables (geographic coordinates and altitude), MLR was developed to estimate the mean, maximum, and minimum average temperature of each month of the year for each location. The average temperatures were calculated as follows (Equation 1):

$$Y_i = \beta_0 + \beta_1 ALT + \beta_2 LAT + \beta_3 LON. \quad (1)$$

where  $Y_i$  is Tmean, Tmax, or Tmin in °C and is the dependent variable. ALT represents the altitude in m, LAT represents the latitude in degrees, and LON represents the longitude in degrees, which are independent variables.  $\beta_0$ ,  $\beta_1$ ,  $\beta_2$ , and  $\beta_3$ , are the regression coefficients. MLR was implemented using the data analysis tool in Microsoft Excel®. Contrary to the methodology applied for ANN and RF, the month was not used as an input variable. Therefore, the data for Tmean, Tmax, and Tmin were classified based on the month. Subsequently, the MLRs were adjusted. Each month had a characteristic equation generating a specific statistical result. The methodology reported by Sediya and Melo Júnior (1998) were used for the studies. This methodology increases the predictive capacity of MLR and facilitates the analysis of each independent variable in the month. The influence of each variable on the result can also be analyzed.

### Artificial Neural Networks (ANNs) model development

ANN was implemented using the Waikato Environment for Knowledge Analysis (WEKA; version 3.8.2 © 1999–2017) developed by the University of Waikato, Hamilton, New Zealand. The algorithm used for ANN was the Multilayer Perceptron (MLP) algorithm (Fausett, 1994). The architecture consisted of the input layer, hidden layers (where the data are processed), and output layer (where the results of processing are compiled) (Figure 2).



**Figure 1:** Climate zoning in the state of Minas Gerais. Zoned according to the Köppen and Geiger (1928) climatic classification. Codes of the climatological stations of the National Institute of Meteorology. Source: Adapted from De Sá Júnior et al. (2012).

**Table 1:** Principal climatological station of the INMET used to estimate the mean, maximum, and minimum air temperature.

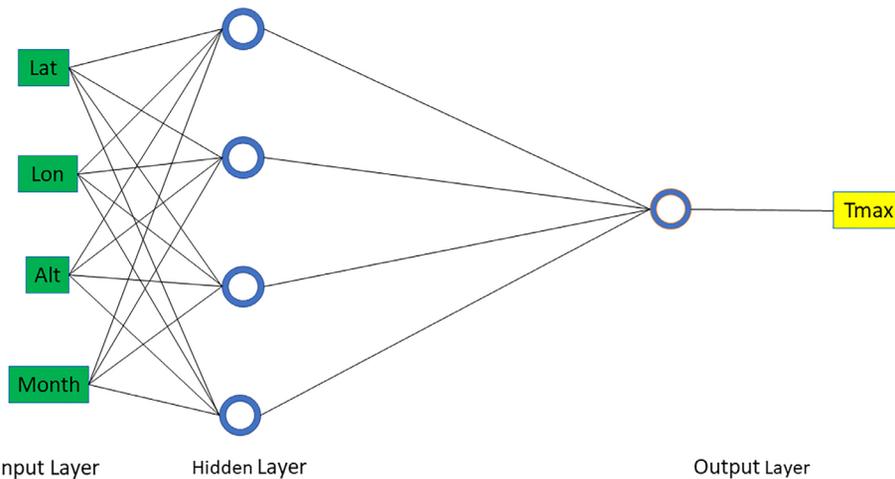
ID	Climatological station location	Latitude (S°)	Longitude (W°)	Altitude (m)	Climatic classes	Period (years)
1	Araxá	19.56	46.93	1004	Cwa	1887 - 2017
2	BambuÍ	20.00	45.98	661	Cwa	1887 - 2017
3	Barbacena	21.25	43.76	1126	Cwb	1890 - 2017
4	Bom Despacho	19.68	45.36	695	Cwa	1887 - 2017
5	Caparaó	20.51	41.86	843	Cwb	1890 - 2017
6	Caratinga	19.80	42.15	609	Cwa	1887 - 2017
7	Coronel Pacheco	21.58	43.25	453	Cwa	1887 - 2009
8	Conceição do Mato Dentro	19.03	43.43	652	Cwa	1887 - 2017
9	Diamantina	18.25	43.60	1296	Cwb	1887 - 2017
10	Florestal	19.88	44.41	760	Cwa	1887 - 2017

Continue...

**Table 1:** Continuation.

11	Ibirité	20.01	44.05	815	Cwa	1887 - 2015
12	Itamarandiba	17.85	42.85	1097	Cwb	1887 - 2017
13	Juiz de Fora	21.76	43.35	940	Cwa	1887 - 2017
14	Lavras	21.23	45.00	919	Cwa	1888 - 2017
15	Machado	21.66	45.91	874	Cwa	1891 - 2017
16	Poços de Caldas	21.91	46.38	1150	Cwb	1892 - 2015
17	São Lourenço	22.10	45.01	900	Cwa	1887 - 2017
18	São Sebastião do Paraíso	20.91	47.11	820	Cwb	1887 - 2013
19	Sete Lagoas	19.46	44.25	732	Cwa	1892 - 2015
20	Viçosa	20.75	42.85	690	Cwa	1890 - 2017

Source: Adapted from De Sá Júnior et al. (2012).



**Figure 2:** Network structure scheme consisting of five neurons in the hidden layer built by WEKA (ANN2) to estimate Tmax.

The input data consisted of the month, latitude, longitude, and altitude of each evaluated location. Each ANN setting estimated the Tmean, Tmax, or Tmin for all the months. There are good reasons behind using these variables for these studies. The temporal variable consists of the cumulative month component, which is required to execute the projections. The latitude and longitude are the variables related to the position. The temperature changes with the position as the position changes from the Poles to the Equator Line. The temperature gradually increases from the poles to the equator. The altitude variable is regarded as the surface component. It can be stated that the higher the altitude, the lower the temperature. The ANN follows a mathematical structure connecting the processing nodes (neurons). The output of a neuron is the

input of the subsequently combined neurons. The final model is built based on various assumptions on activation function (Equations 2, 3, 4, 5, 6, 7 and 8). The equations are as follows:

$$\overline{N}_1 = L n(1 + e^{(w_{1,1}LAT + w_{2,1}LON + w_{3,1}ALT + w_{4,1}MONTH + w_{5,1}B)}) \quad (2)$$

$$\overline{N}_2 = L n(1 + e^{(w_{1,2}LAT + w_{2,2}LON + w_{3,2}ALT + w_{4,2}MONTH + w_{5,2}B)}) \quad (3)$$

$$\overline{N}_3 = L n(1 + e^{(w_{1,3}LAT + w_{2,3}LON + w_{3,3}ALT + w_{4,3}MONTH + w_{5,3}B)}) \quad (4)$$

$$\overline{N}_4 = L n(1 + e^{(w_{1,4}LAT + w_{2,4}LON + w_{3,4}ALT + w_{4,4}MONTH + w_{5,4}B)}) \quad (5)$$

$$\overline{T}_{max} = L n(1 + e^{(w_{1,5}N_1 + w_{2,5}N_2 + w_{3,5}N_3 + w_{4,5}N_4 + w_{5,5}B)}) \quad (6)$$

$$\overline{T}_{\min} = L n(1 + e^{(w_1, s^{N_1} + w_2, s^{N_2} + w_3, s^{N_3} + w_4, s^{N_4} + w_5, s^{\beta})}) \quad (7)$$

$$\overline{T}_{\text{mean}} = L n(1 + e^{(w_1, s^{N_1} + w_2, s^{N_2} + w_3, s^{N_3} + w_4, s^{N_4} + w_5, s^{\beta})}) \quad (8)$$

Equations 2–5 represent the mathematical abstraction of the ANN built in Figure 2 extracting the neurons equations. Equations 6 – 8 are the estimate vectors of each output.  $W_{ij}$  represents the weights estimated using the backpropagation algorithm during ANN processing. The value of  $B_{ij}$  represents the bias associated with each measurer. The activation function applied was sigmoidal with non-linear output.

All adjustments were cross-assessed. Twenty folds of the sample set were used for the assessment for training to compensate for the reduced number of instances. Two different configurations were evaluated (Table 2). Results from the preliminary tests indicated that changes in the number of training epochs and the number of neurons present in the hidden layer interfered with the performance of the models. However, changes

in the other parameters did not significantly influence the model performances.

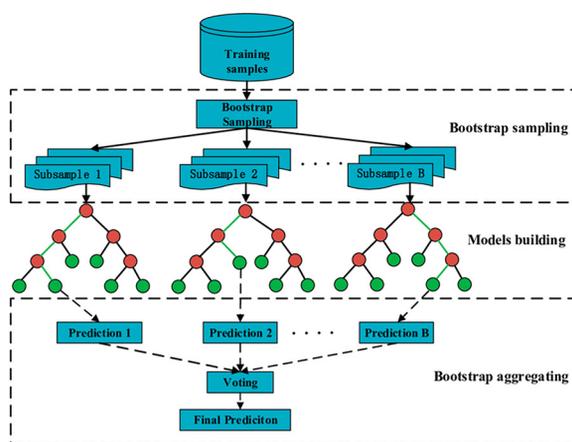
### Development of the Random Forest (RF) model

The implementation of RF in WEKA has its basis on a previously reported study (Breiman, 2001). Two configurations of RF were used, with the input variables being month, latitude, longitude, and altitude of each evaluated location. Thus, each RF setting could be used to estimate Tmean, Tmax, or Tmin for all the months under study. The steps followed has been presented in Figure 3.

In this study, preliminary examinations were conducted for several configurations. The configurations with 100 and 500 interactions exhibited better performance compared to other values obtained in the preliminary analysis. The preliminary tests revealed that the changes in the other parameters did not positively influence the model performance. The tests exhibited two distinct configurations for better results (Table 3).

**Table 2:** WEKA configuration in the ANN implementation.

	Tmean		Tmax		Tmin	
	ANN1	ANN2	ANN1	ANN2	ANN1	ANN2
Learning rate	0.3	0.3	0.3	0.3	0.3	0.3
Momentum	0.2	0.2	0.2	0.2	0.2	0.2
Number of training epochs	500	1000	500	500	500	1000
Number of hidden layers	1	1	1	1	1	1
Number of neurons into the hidden layer	5	6	6	5	6	6



**Figure 3:** Schematic representation of the steps used in the RF model following the resampling strategy (Source: Wang et al., 2019).

**Table 3:** WEKA configuration in the RF implementation process.

	Tmean		Tmax		Tmin	
	RF1	RF2	RF1	RF2	RF1	RF2
Break ties randomly when several attributes look equally good		x		x	x	x
Size of each bag, as percentage of the training set size	100	100	100	100	100	100
Number of iterations	500	500	500	500	100	500
Minimum number of instances per leaf	1	1	1	1	1	1
Maximum depth of the tree	unl.	unl.	unl.	unl.	unl.	unl.

### Statistical tests

Various statistical indices were used to assess the predictive quality of each technique in terms of variation, precision, accuracy, and performance. The mean absolute error (MAE) and root mean square error (RMSE) indicates revealed how close the predicted values were to the observed value. Thus, the accuracy of each model could be predicted. The variation was quantified by the determination coefficient ( $R^2$ ), which represents the percentage of the variation of the dependent variable explained by the independent variable. The best model should produce an  $R^2$  value close to unity. The precision of the models was quantified based on Pearson's correlation coefficient ( $r$ ), which indicates the degree of dispersion of the data obtained in terms of the mean. Accuracy was quantified using Willmott's index of agreement ( $d$ ) and the performance index ( $c$ ) (Camargo; Sentelhas, 1997). The performance index was calculated using the equation  $c = r \cdot d$ . This equation was also used to quantify the performance of the model. The performances were classified as: Excellent (1 – 0.85), Very good (0.85 – 0.76), Good (0.76 – 0.66), Average (0.66 – 0.61), Poor (0.61 – 0.51), Bad (0.51 – 0.41), and Terrible (less than 0.41).

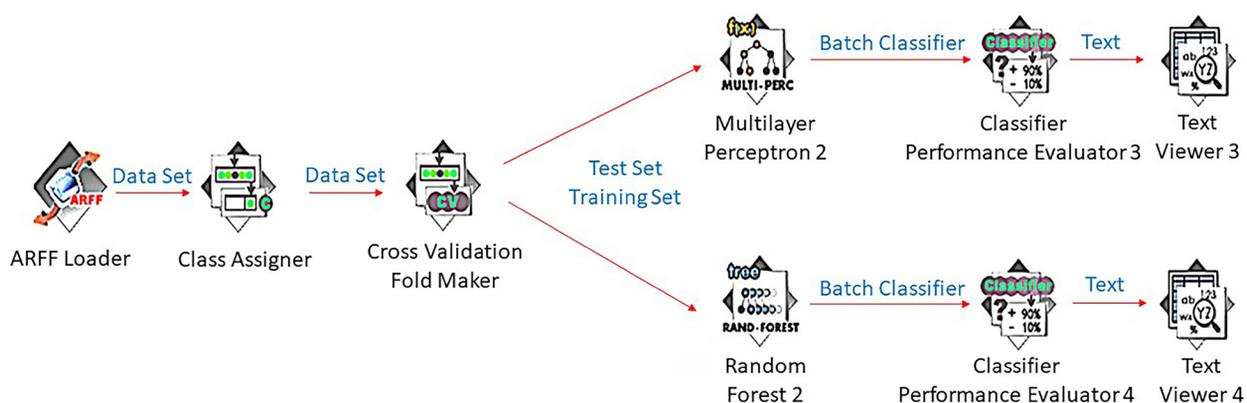
Weka provides a tool to compare different combinations and different algorithms called WEKA Experiment Environment (Figure 4). This tool was used to compare the performance of each algorithm and configuration used in the present study conducted using the cross-validation technique. According to Noi, Degener, and Kappas, (2017), cross-validation is one of the most popular validation methods used to compare different combinations and different algorithms. In the cross-validation method, the dataset is divided into  $k$  groups ( $k$ -fold) of approximately the same size. Due to the number of observations, a 20-fold cross-validation method was used. The algorithms were applied for each fold, generating statistical performance values. Later, these average

performance values were compared by Tukey's test at 5% probability. The statistical software Sisvar (Ferreira, 2019) was used for analysis. The MLR method was not implemented in WEKA. The approach was different from that was used in the ANN and RF methods. Hence, it was not possible to compare the MLR method with the other techniques using Tukey's test. The comparison between MLR and other techniques was made by comparing the statistical performance indicators.

## RESULTS AND DISCUSSION

The MLR method coefficients were adjusted to estimate the Tmean, Tmax, and Tmin monthly air temperatures. The respective mean absolute errors (MAE), root mean square errors (RMSE), determination coefficient ( $R^2$ ), Pearson's correlation coefficient ( $r$ ), Willmott's index of agreement ( $d$ ), and the consistency index ( $c$ ) are shown in Table 4.

The models used to estimate Tmean (Table 4) reveal that  $R^2$  values were in the range of 0.38 – 0.93 and the  $r$  valued ranged from 0.62 to 0.97. The models for estimating the data for the months of July and August exhibited a “bad” and “poor” performance (Camargo; Sentelhas, 1997), respectively. For these months, these models are not recommended to estimate the Tmean values. The model performances were “Good” when the other months were analyzed. The linear coefficients altitude ( $\beta_1$ ) and latitude ( $\beta_2$ ) were significant. A negative correlation was observed between altitude and Tmean and between latitude and Tmean, exhibiting a decrease in Tmean values with increasing altitude and latitude. These results were expected and in accordance with the vertical thermal gradient in the troposphere. Cargnelutti Filho, Maluf and Matzenauer (2008) and Gomes et al. (2014) reported a negative correlation between altitude and Tmean (Rio de Janeiro state and the Rio Grande do Sul state, respectively). However, there was no significant influence in latitude.



**Figure 4:** WEKA Experiment Environment workflow of the experiment.

**Table 4:** Coefficients of the monthly air temperature models and statistical performance indicators.

Month	MLR method coefficients				MAE (°C)	RMSE (°C)	R <sup>2</sup>	r	d	c
	( $\beta_0$ )	Alt ( $\beta_1$ )	Lat ( $\beta_2$ )	Lon ( $\beta_3$ )						
Tmean										
Jan	26.18	0.2980*	-0.144*	0.004*	0.25	0.33	0.88	0.94	0.97	0.90
Feb	26.41	-0.0030*	-0.346*	0.143 <sup>ns</sup>	0.33	0.41	0.75	0.86	0.92	0.80
Mar	27.05	-0.0040*	-0.311*	0.114 <sup>ns</sup>	0.28	0.34	0.85	0.92	0.96	0.89
Apr	22.87	-0.0046*	-0.363*	0.214*	0.38	0.45	0.82	0.90	0.95	0.86
May	28.84	-0.0026*	-0.760*	0.162 <sup>ns</sup>	0.47	0.56	0.75	0.87	0.92	0.80
Jun	33.19	-0.0011 <sup>ns</sup>	-0.917*	0.080 <sup>ns</sup>	0.62	0.72	0.65	0.80	0.88	0.71
Jul	25.62	-0.0011 <sup>ns</sup>	-0.718*	0.156 <sup>ns</sup>	0.86	0.99	0.38	0.62	0.73	0.45
Aug	21.73	-0.0014 <sup>ns</sup>	-0.779*	0.307 <sup>ns</sup>	0.75	0.87	0.51	0.71	0.81	0.58
Sep	17.94	-0.0028*	-0.828*	0.491*	0.63	0.72	0.71	0.84	0.91	0.77
Oct	18.72	-0.0046*	-0.613*	0.442*	0.44	0.49	0.85	0.92	0.96	0.88
Nov	18.29	-0.0061*	-0.409*	0.387*	0.27	0.32	0.93	0.97	0.98	0.95
Dec	21.63	-0.0053*	-0.173*	0.205*	0.23	0.30	0.91	0.95	0.98	0.93
Tmax										
Jan	35.07	-0.0051*	-0.468*	0.179 <sup>ns</sup>	0.52	0.61	0.76	0.87	0.93	0.81
Feb	36.87	-0.0037*	-0.483*	0.141 <sup>ns</sup>	0.60	0.70	0.63	0.80	0.88	0.70
Mar	31.98	-0.0058*	-0.332*	0.199 <sup>ns</sup>	0.46	0.51	0.83	0.91	0.95	0.87
Apr	25.54	-0.0061*	-0.511*	0.407*	0.50	0.58	0.83	0.91	0.95	0.87
May	25.88	-0.0055*	-0.885*	0.509*	0.49	0.58	0.86	0.93	0.96	0.89
Jun	23.46	-0.0068*	-0.753*	0.514*	0.62	0.69	0.83	0.91	0.95	0.87
Jul	20.14	-0.0059*	-0.747*	0.577*	0.55	0.66	0.82	0.91	0.95	0.86
Aug	13.30	-0.0067*	-0.704*	0.762*	0.53	0.64	0.86	0.93	0.96	0.89
Sep	15.71	-0.0054*	-0.992*	0.855*	0.64	0.74	0.84	0.91	0.95	0.87

Continue.

**Table 4:** Continuation.

Month	MLR method coefficients				MAE (°C)	RMSE (°C)	R <sup>2</sup>	r	d	c
	( $\beta_0$ )	Alt ( $\beta_1$ )	Lat ( $\beta_2$ )	Lon ( $\beta_3$ )						
Oct	15.63	-0.0069*	-0.690*	0.753*	0.61	0.69	0.84	0.92	0.95	0.88
Nov	17.02	-0.0071*	-0.480*	0.609*	0.52	0.60	0.85	0.92	0.96	0.88
Dec	22.13	-0.0071*	-0.101*	0.336*	0.45	0.51	0.84	0.92	0.96	0.88
Tmin										
Jan	19.08	-0.0038*	-0.071 <sup>ns</sup>	0.090 <sup>ns</sup>	0.35	0.42	0.71	0.84	0.91	0.77
Feb	22.79	-0.0025*	-0.188 <sup>ns</sup>	0.031 <sup>ns</sup>	0.49	0.56	0.47	0.69	0.80	0.55
Mar	25.67	-0.0023 <sup>ns</sup>	-0.241 <sup>ns</sup>	0.019 <sup>ns</sup>	0.51	0.66	0.43	0.66	0.78	0.51
Apr	30.60	-0.0032 <sup>ns</sup>	-0.222 <sup>ns</sup>	0.164 <sup>ns</sup>	0.77	0.97	0.41	0.64	0.78	0.50
May	39.19	-0.0005 <sup>ns</sup>	-0.690 <sup>ns</sup>	0.262 <sup>ns</sup>	0.92	1.16	0.37	0.60	0.73	0.44
Jun	40.56	0.0004 <sup>ns</sup>	-0.653 <sup>ns</sup>	0.362 <sup>ns</sup>	1.17	1.49	0.25	0.50	0.62	0.31
Jul	33.88	0.0015 <sup>ns</sup>	-0.577 <sup>ns</sup>	0.287 <sup>ns</sup>	1.52	1.92	0.10	0.32	0.40	0.13
Aug	30.86	0.0001 <sup>ns</sup>	-0.578 <sup>ns</sup>	0.173 <sup>ns</sup>	1.43	1.71	0.12	0.35	0.47	0.17
Sep	30.06	-0.0002 <sup>ns</sup>	-0.733 <sup>ns</sup>	0.021 <sup>ns</sup>	1.20	1.37	0.22	0.47	0.61	0.29
Oct	26.21	-0.0028 <sup>ns</sup>	-0.506 <sup>ns</sup>	0.067 <sup>ns</sup>	0.66	0.81	0.50	0.71	0.81	0.57
Nov	23.23	-0.0049*	-0.344 <sup>ns</sup>	0.126 <sup>ns</sup>	0.31	0.40	0.86	0.93	0.96	0.89
Dec	22.09	-0.0039*	-0.133 <sup>ns</sup>	0.054 <sup>ns</sup>	0.29	0.36	0.82	0.91	0.94	0.86

<sup>ns</sup> no significant. \*significant at 5% probability by F-test.

During the estimation of Tmax, RMSE was found to be in the range of 0.51 – 0.74. The R<sup>2</sup> values ranged between 0.63 and 0.86, and the r values ranged between 0.80 and 0.93 (Table 4). The model for February exhibited the lowest statistical indicators, and the model's performance was "Good" (Camargo; Sentelhas, 1997). The linear coefficient of altitude ( $\beta_1$ ) was significant in all models. There was no significant influence of the linear coefficients longitude ( $\beta_3$ ) on the months of January, February, and March. In the other months, a significant influence of  $\beta_2$ ,  $\beta_3$ , and  $\beta_4$  was observed. Gomes et al. (2014) analyzed the models to estimate the maximum monthly air temperature of Rio de Janeiro. R<sup>2</sup> values were found to be in the range of 0.51 – 0.71. A significant influence of the altitude and latitude was observed. However, the linear coefficient of longitude did not significantly affect the data of most months. This difference can be explained by the small longitudinal difference between the meteorological stations in Rio de Janeiro state compared to the region evaluated in this study. The meteorological stations under consideration are at a sufficient longitudinal distance to be influenced by the continentality effect.

While estimating Tmin, it was observed that the r values ranged between 0.32 and 0.93. The R<sup>2</sup> values

ranged between 0.10 and 0.86, and the RMSE values ranged between 0.36 – 1.92 (Table 4). The models used for estimating the Tmin values for the months between February and October exhibited a "Poor", "Bad", or "Terrible" performance index (Camargo; Sentelhas, 1997), reflecting the low precision and degree of accuracy. Furthermore, significant  $\beta_1$ ,  $\beta_2$ , and  $\beta_3$  values were not recorded when these models were used to study the data corresponding to the abovementioned months.

The Tmin, corresponding to these months, varied due to the variation in other factors, such as wind, ocean currents, local topographic conditions, rain, cloudiness, and passage of the cold front (Aguado; Burt, 2010). According to Silveira et al. (2019), in addition to the statistical factors (vegetation, maritime, continentality, geographic coordinates, etc.), climatic conditions are influenced by dynamic atmospheric systems such as cold fronts. After the passage of the cold front, under conditions of clear skies and low atmospheric humidity, the heat loss by irradiation during the night is very high. This results in a drop in temperature, mainly during winter, autumn, and spring. In some cases, this facilitates the occurrence of radioactive frosts (Escobar, 2007).

Therefore, the  $T_{min}$  values could not be estimated with high precision using these models. In the other months (November, December, and January), the models performed well, and a significant influence of altitude was observed. Medeiros et al. (2005) (the Northeast region of Brazil) Cargnelutti Filho et al. (2006) (the Rio Grande do Sul state), and Gomes et al. (2014) (the Rio de Janeiro state), observed similar results. The altitude influenced the  $T_{min}$  values the most.

The ANN and RF statistical performance indicators for estimating  $T_{mean}$ ,  $T_{max}$ , and  $T_{min}$  in the regions classified as Cwa and Cwb (Minas Gerais state) are shown in Table 5. Contrary to the MLR model, which used separate equations for each month, the architectures chosen for the ANN and RF models could be used to estimate the  $T_{mean}$ ,  $T_{max}$ , and  $T_{min}$  of all months together. Thus, to estimate the  $T_{mean}$ ,  $T_{max}$ , or  $T_{min}$  of a given location, latitude, longitude, altitude, and the month were used as the input data. Moreover, the statistics for each configuration (Table 5) refer to all the months of the year. The model performance indices for each month need not be distinguished (unlike the MLR model).

**Table 5:** Summary of the statistical tests conducted using the ANN and RF models.

	$T_{mean}$					
	MAE	RMSE	R <sup>2</sup>	r	d	c
RF1	0.47 a	0.61 a	0.94	0.97 a	0.98	0.95
RF2	0.43 a	0.57 a	0.95	0.98 a	0.99	0.96
ANN1	0.67 b	0.80 b	0.90	0.96 a	0.97	0.92
ANN2	0.64 b	0.78 b	0.91	0.96 a	0.98	0.93
	$T_{max}$					
	MAE	RMSE	R <sup>2</sup>	r	d	c
RF1	0.46 a	0.55 a	0.94	0.97 a	0.98	0.95
RF2	0.44 a	0.56 a	0.94	0.97 a	0.98	0.95
ANN1	0.73 b	0.86 b	0.85	0.93 b	0.96	0.88
ANN2	0.65 b	0.79 b	0.87	0.94 b	0.97	0.90
	$T_{min}$					
	MAE	RMSE	R <sup>2</sup>	r	d	c
RF1	0.59 a	0.77 a	0.94	0.96 a	0.98	0.95
RF2	0.58 a	0.76 a	0.94	0.97 a	0.98	0.95
ANN1	0.88 b	1.07 b	0.89	0.94 a	0.97	0.91
ANN2	0.87 b	1.03 b	0.89	0.94 a	0.97	0.92

Mean values followed by the same letter in a column do not differ significantly (Tukey's test ( $p \leq 0.05$ )).

The lower RMSE and MAE were observed when the RF technique was used (compared to the case when ANN was used). A significant difference was observed in the results obtained using these techniques (ANN and RF). There was no significant difference between the different configurations tested within each technique. The RMSE and MAE were higher estimating  $T_{min}$  values compared to the  $T_{max}$  and  $T_{mean}$  values, suggesting more variation within the  $T_{min}$  estimates. The  $r$  values, calculated using the RF method, were higher than those calculated using the ANN method during the calculation of the  $T_{mean}$ ,  $T_{max}$ , and  $T_{min}$  values. The values of the coefficient  $r$  did not differ significantly when these two techniques (and different configurations of the techniques) were used to determine the  $T_{mean}$  and  $T_{min}$  values. Nevertheless, a significant difference was observed in the  $T_{max}$  values when these two techniques were used. The other indices indicate that the RF model was superior to the ANN model. However, both the techniques could be used to estimate the  $T_{mean}$ ,  $T_{max}$ , and  $T_{min}$  values with very high accuracy (Table 5). The fit quality of both models can be confirmed by the high values of the performance index ( $c$ ). These values were "Excellent" according to the evaluation criteria proposed by Camargo and Sentelhas (1997).

There was no significant difference between the RF configurations. However, the use of the concept of *Break ties randomly when several attributes look equally good*, a WEKA solution, increased the predictive capacity of the model. That option gets triggered when the output reaches a local optimum. When this condition becomes true, the algorithm initializes a random process to escape from a local optimal spot to reach the best solutions. This procedure has been explained in detail by Breiman (2001). The previous studies suggest the execution of 100 interactions; however, 500 interactions were required to improve the RF performance. Were et al. (2015) reported more stable results using a higher number of interactions.

The changes made to the ANN parameters did not significantly influence the  $T_{mean}$ ,  $T_{max}$ , and  $T_{min}$  values. However, increasing the *Number of Training Epochs* from 500 to 1000 improved the  $T_{mean}$  and  $T_{min}$  predictive capacity of ANN. This *Number of Training Epochs* is a hyperparameter that defines the number of times the learning algorithm works through the entire training dataset. The best results were obtained when six neurons were integrated into one hidden layer during the estimation of  $T_{mean}$  and  $T_{min}$ . However, the best result was obtained when five neurons were integrated into the hidden layer during the estimation of  $T_{max}$ . The choice of the size of

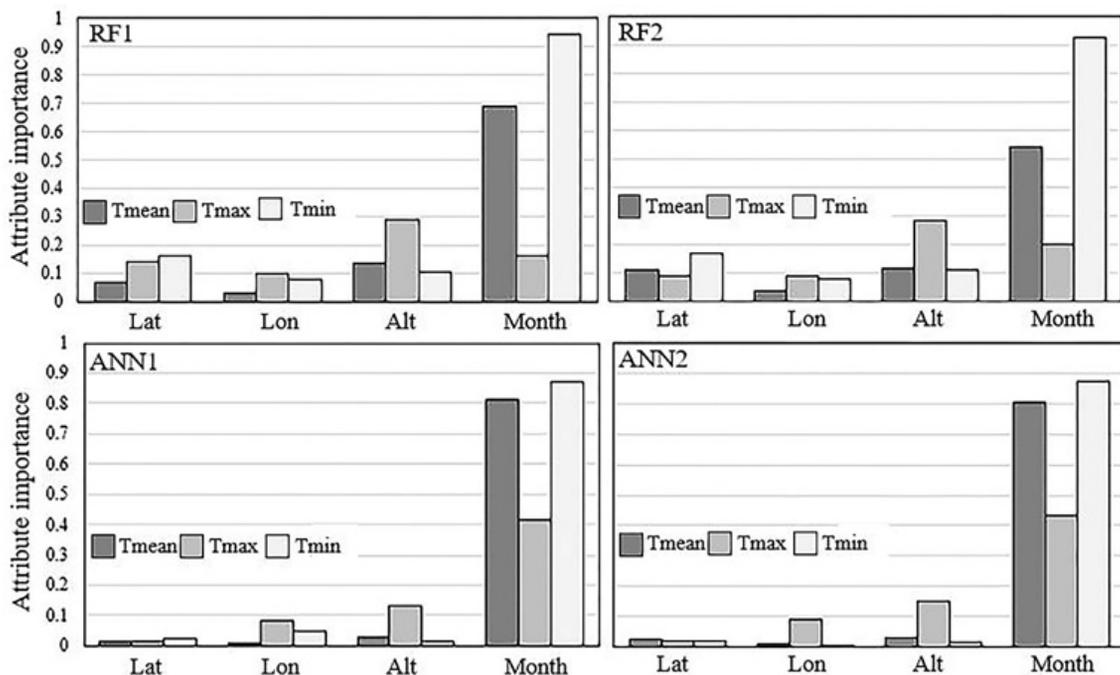
the hidden layer is very important because underestimated numbers of neurons can lead to poor approximation and generalization capabilities, while the use of excessive neurons can potentially result in overfitting. This can eventually make the search for the global optimum more difficult (Lee; Lam, 1995).

Although the MLR model could be used to estimate the Tmean, Tmax, and Tmin for some months of the year, in general, the RF and ANN models exhibited superior predictive abilities (for all the analyzed statistical indices) than the MLR models. The RF model was found to be superior to the ANN model. Moreover, the low MLR predictive capacity (Tmin estimation) can cause problems for producers who need this information because the regions categorized as Cwa and Cwb are more suitable for the development of agricultural activities that require lower temperatures and average temperatures during the winter (below 20 °C; De Sá Júnior et al., 2012). Therefore, RF and ANN methods are more suitable for this region.

Several literature reports (reporting various applications) have indicated the superiority of the RF model in the regression estimation (Benali et al., 2019; Noi; Degener; Kappas, 2017; Rodríguez-Lado et al., 2015). The superiority of the RF model can be attributed to the advantages of the method, which include not making distributive assumptions about the predictors.

The importance of each variable can be determined using this model, and the method is less sensitive to noise or overfitting (Armitage; Ober, 2010; Ismail; Mutanga, 2010). Even though RF is superior to ANN, the ANN method can be used to determine the Tmean, Tmax, and Tmin values with high accuracy. This has also been reported by Hasni et al. (2012). They concluded that the ANN technique could be reliably used for determining the temperatures.

The plot, shown in Figure 5, indicates the importance of each input attribute in the response variable of the evaluated algorithms. The most important contribution toward the estimation of the Tmean value was for the month. This was followed by the effect of the altitude (for all the evaluated models). In the estimate of Tmax by RF1 and RF2, the altitude exerted the maximum effect. However, when the ANN1 and ANN2 methods were used, the month was found to exert the maximum effect on the results. This was followed by the contribution of the altitude. The trend was similar to the trend observed when the MLR method was used. A significant influence of the altitude was observed for all months when the MLR model was used for the calculations. The month attribute had the largest contribution to the Tmin estimate. This contribution was the maximum. These results can potentially explain the low capacity of the MLR model toward the estimation of Tmin as the month is not considered a variable in this model.



**Figure 5:** Attribute importance plots for the RF1, RF2, ANN1, and ANN2 models (Source: The Authors).

The results revealed that, for locations where it is difficult to collect data from weather stations (due to lack of infrastructure, reading errors, or use of damaged devices), the use of RF and ANN models is recommended for estimating the Tmean, Tmax, and Tmin values. In addition, researchers and producers can use such methods to create a risk zoning of pests and diseases, develop works related to plant growth, and develop crop varieties based on the temperature of the region.

An estimation of the Tmin values can help prevent the formation of frost in all the locations under study. This is because the region under analysis is susceptible to the occurrence of this phenomenon. According to Pimenta, Angélico and Chalfoun (2018), adverse weather conditions (such as the formation of frost) can harm the production of the coffee fruit, affecting productivity and thereby changing the market value of the product. It is important to develop an efficient technique to determine the Tmax and Tmin values to develop a more accurate agricultural zoning of climatic risk. This can assist the producers in the choice of sowing time and harvest planning. Extreme weather conditions, especially in less developed regions, can be avoided. However, no statistical method can produce results that are exactly the same as the observed and/or recorded data. Hence, it is important that the weather stations function continuously (Alves et al., 2020). Furthermore, it is important to have computational knowledge to implement the RF and ANN models, therefore, mobile applications are needed to facilitate the use of these techniques. Further studies in the area are needed, and the results of the present study may support future forecasts.

## CONCLUSIONS

The results of this study can help farmers, researchers, technicians, and local government officials in urban planning. Urbanization is characterized by surface alterations. Vegetated areas are replaced with impervious surfaces and buildings. This surface change alters the energy balance, increasing absorption and heat transfer between the earth's surface and the lower atmosphere, resulting in increased surface air temperatures (Song; Wu, 2016). Accelerated urban growth has been observed in the region under study. An effective tool for estimating the air temperature can assist in the application of new technologies that can potentially reduce the surface heating process. The RF model exhibited a greater predictive performance compared to the ANN and MLR models for estimating the Tmean, Tmax, and Tmin values. The RF model explains at least 94% of

the variability of the variables estimated using the independent dataset, i.e., only 6% of the response variable could not be predicted by the model. The RF is the most suitable technique for estimating the air temperature. The input attributes were sufficient for the estimation. Therefore, this model is recommended for conducting studies in this specific region.

## ACKNOWLEDGEMENTS

The authors express their gratitude to CAPES for scholarships that enabled the development of this research, and Brazilian National Institute of Meteorology (INMET) for making the series of meteorological data available.

## REFERENCES

- AGUADO, E.; BURT, J. E. **Understanding weather and climate**. 5<sup>th</sup> ed. New Jersey: Prentice Hall. 2010. 505p.
- ALVARES, C. A. et al. Modeling monthly mean air temperature for Brazil. **Theoretical and Applied Climatology**, 113(3-4):407-427, 2013.
- ALVES, M. P. A. et al. Reconstrução de dados e detecção de ondas de calor e de frio no Porto e concelhos vizinhos-Portugal. **Territorium**, 27(II):49-66, 2020.
- ANTONOPOULOS, V. Z.; ANTONOPOULOS, A. V. Daily reference evapotranspiration estimates by artificial neural networks technique and empirical equations using limited input climate variables. **Computers and Electronics in Agriculture**, 132:86-96, 2017.
- ARMITAGE, D. W.; OBER, H. K. A comparison of supervised learning techniques in the classification of bat echolocation calls. **Ecological Informatics**, 5(6):465-473, 2010.
- ASADI, H. et al. Rainfall-runoff modelling using hydrological connectivity index and artificial neural network approach. **Water**, 11(2):1-20, 2019.
- BATISTA-SANTOS, P. et al. The impact of cold on photosynthesis in genotypes of *Coffea* spp.:Photosystem sensitivity, photoprotective mechanisms and gene expression. **Journal of Plant Physiology**, 168(8):792-806, 2011.
- BENALI, L. et al. Solar radiation forecasting using artificial neural network and random forest methods: Application to normal beam, horizontal diffuse and global components. **Renewable Energy**, 132:871-884, 2019.
- BENAVIDES, R. et al. Geostatistical modelling of air temperature in a mountainous region of Northern Spain. **Agricultural and Forest Meteorology**, 146(3-4):173-188, 2007.

- BENLLOCH-GONZÁLEZ, M. et al. Effect of moderate high temperature on the vegetative growth and potassium allocation in olive plants. **Journal of Plant Physiology**, 207:22-29, 2016.
- BOU-RABEE, M. et al. Using artificial neural networks to estimate solar radiation in Kuwait. **Renewable and Sustainable Energy Reviews**, 72:434-438, 2017.
- BRASIL. Ministério da Agricultura e Reforma Agrária. Secretaria Nacional de Irrigação. Departamento Nacional de Meteorologia. **Normais climatológicas (1961-1990)**. Brasília: 1992. 84p.
- BREIMAN, L. Random forests. **Machine Learning**, 45(1):5-32, 2001.
- CAMARGO, A. P. de.; SENTELHAS, P. C. Performance evaluation of different potential evapotranspiration estimating methods in the State of São Paulo, Brazil. **Revista Brasileira de Agrometeorologia**, 5(1):89-97, 1997.
- CANNELL, M. G. R. Physiology of the coffee crop. In: CLIFFORD, M. N.; WILLSON, K. C. (eds). **Coffee**: Boston. MA: Springer, p.108-134, 1985.
- CARDOSO, M. R. D. et al. Caracterização da temperatura do ar no estado de Goiás e no Distrito Federal. **Revista Brasileira de Climatologia**, 11:119-134, 2012.
- CARDOSO, M. R. D. et al. Classificação climática de Köppen-Geiger para o estado de Goiás e o Distrito Federal. **Acta Geográfica**, 8(16):40-55, 2015.
- CARGNELUTTI FILHO, A.; MALUF, J. R. T.; MATZENAUER, R. Coordenadas geográficas na estimativa das temperaturas máxima e média decendiais do ar no Estado do Rio Grande do Sul. **Ciencia Rural**, 38(9):2448-2456, 2008.
- CARGNELUTTI FILHO, A. et al. Altitude e coordenadas geográficas na estimativa da temperatura mínima média decendial do ar no Estado do Rio Grande do Sul. **Pesquisa Agropecuária Brasileira**, 41(6):893-901, 2006.
- COMPANHIA BRASILEIRA DE ABASTECIMENTO - CONAB. **Acompanhamento da safra brasileira de café**. Safra 2020 - Primeiro Levantamento. 6:1-62. 2020. Available in: <<https://www.conab.gov.br/info-agro/safras/cafe>>. Access in: April, 28, 2020.
- COSTA, H. C. et al. Espacialização e sazonalidade da precipitação pluviométrica do estado de Goiás e Distrito Federal. **Revista Brasileira de Geografia Física**, 1:87-100, 2012.
- DAMATTA, F. M. et al. Physiological and agronomic performance of the coffee crop in the context of climate change and global warming: A review. **Journal of Agricultural and Food Chemistry**, 66(21):5264-5274, 2018.
- DE OLIVEIRA APARECIDO, L. E. et al. Machine learning algorithms for forecasting the incidence of *Coffea arabica* pests and diseases. **International Journal of Biometeorology**, 64:671-688, 2020.
- DE SÁ JÚNIOR, A. et al. Application of the Köppen classification for climatic zoning in the state of Minas Gerais. Brazil. **Theoretical and Applied Climatology**, 108(1-2):1-7, 2012.
- DUMEDAH, G.; COULIBALY, P. Evaluation of statistical methods for infilling missing values in high-resolution soil moisture data. **Journal of Hydrology**, 400(1-2):95-102, 2011.
- ESCOBAR, G. C. J. Padrões sinóticos associados a ondas de frio na cidade de São Paulo. **Revista Brasileira de Meteorologia**, 22(2):241-254, 2007.
- FABRIS, F. et al. A new approach for interpreting random forest models and its application to the biology of ageing. **Bioinformatics**, 34(14):2449-2456, 2018.
- FAUSETT, L. **Fundamentals of neural networks: Architectures, algorithms and applications**. Prentice Hall. Upper Saddle River. New Jersey. 1994. 461p
- FERREIRA, D. F. SISVAR: A computer analysis system to fixed effects split plot type designs. **Revista Brasileira de Biometria**, 37(4):529-535, 2019.
- GOMES, D. P. et al. Estimativa da temperatura do ar e da evapotranspiração de referência no estado do Rio de Janeiro. **Irriga**, 19(2):302-314, 2014.
- HASNI, A. et al. Estimating global solar radiation using artificial neural network and climate data in the south-western region of Algeria. **Energy Procedia**, 18:531-537, 2012.
- HESTERBERG, T. et al. Bootstrap methods and permutation tests. In: MOORE, D. S. **The Practice of Business Statistics**. NY: W.H. Freeman and Co. Chap.18, p.4-25, 2002.
- ISMAL, R.; MUTANGA, O. A comparison of regression tree ensembles: Predicting *Sirex noctilio* induced water stress in *Pinus patula* forests of KwaZulu-Natal. South Africa. **International Journal of Applied Earth Observation and Geoinformation**, 12(1):S45-S51, 2010.
- JAMES, G. et al. **An introduction to statistical learning**. New York: Springer. 2013. v. 112. 18p.
- KÖPPEN, W.; GEIGER, R. **Klimate der Erde**. Gotha: Verlag Justus Perthes. Wall-Map 150cmx200cm. 1928.

- KUMAR, M.; RAGHUWANSHI, N. S.; SINGH, R. Artificial neural networks approach in evapotranspiration modeling: A review. **Irrigation Science**, 29(1):11-25, 2011.
- LEE, K. W.; LAM, H. N. Optimal sizing of feedforward neural networks: Case studies. **Proceedings 1995 Second New Zealand International Two-Stream Conference on Artificial Neural Networks and Expert Systems**, p.79-82, 1995.
- MEDEIROS, S. S. de. et al. Estimativa e espacialização das temperaturas do ar mínimas, médias e máximas na Região Nordeste do Brasil. **Revista Brasileira de Engenharia Agrícola e Ambiental**, 9(2):247-255, 2005.
- MOREIRA, M. C.; CECÍLIO, R. A. Software to estimate air temperature in the Brazilian northeastern region using artificial neural networks. **Revista Engenharia na Agricultura-Reveng**, 24(2):164-171, 2016.
- MUHAMMAD, W. et al. Pancreatic cancer prediction through an artificial neural network. **Frontiers in Artificial Intelligence**, 2:2, 2019.
- MWALE, F. D.; ADELOYE, A. J.; RUSTUM, R. Infilling of missing rainfall and streamflow data in the Shire River basin, Malawi: A self organizing map approach. **Physics and Chemistry of the Earth**, 50(52):34-43, 2012.
- NOI, P. T.; DEGENER, J.; KAPPAS, M. Comparison of multiple linear regression, cubist regression and random forest algorithms to estimate daily air surface temperature from dynamic combinations of MODIS LST data. **Remote Sensing**, 9(5):398, 2017.
- PARTELLI, F. L. et al. Low temperature impact on photosynthetic parameters of coffee genotypes. **Pesquisa Agropecuária Brasileira**, 44(11):1404-1415, 2009.
- PEZZOPANE, J. et al. Espacialização da temperatura do ar no Estado do Espírito Santo. **Revista Brasileira de Agrometeorologia**, 12(1):151-158, 2004.
- PIMENTA, C. J.; ANGÉLICO, C. L.; CHALFOUN, S. M. Challenges in coffee quality: Cultural. **Ciência e Agrotecnologia**, 42(4):337-349, 2018.
- RODRÍGUEZ-LADO, L. et al. A pedotransfer function to map soil bulk density from limited data. **Procedia Environmental Sciences**, 27:45-48, 2015.
- SCORNET, E. On the asymptotics of random forests. **Journal of Multivariate Analysis**, 146:72-83, 2016.
- SEDIYAMA, G. C.; MELO JÚNIOR, J. C. F. Modelos para estimativa das temperaturas normais mensais médias máximas, mínimas e anual no estado de Minas Gerais. **Engenharia na Agricultura**, 6(1):57-61, 1998.
- SEYEDHOSSEINI, M.; TASDIZEN, T. Disjunctive normal random forests. **Pattern Recognition**, 48(3):976-983, 2015.
- SILVEIRA, R. B. et al. Ondas de calor nas capitais do Sul do Brasil e Montevideu - Uruguai. **Revista Brasileira de Geografia Física**, 12(4):1259-1276, 2019.
- SONG, Y.; WU, C. Examining the impact of urban biophysical composition and neighboring environment on surface urban heat island effect. **Advances in Space Research**, 57(1):96-109, 2016.
- VOGELS, M. F. A. et al. Agricultural cropland mapping using black-and-white aerial photography object-based image analysis and random forests. **International Journal of Applied Earth Observation and Geoinformation**, 54:114-123, 2017.
- WAHID, A. et al. Heat tolerance in plants: An overview. **Environmental and Experimental Botany**, 61(3):199-223, 2007.
- WANG, H. et al. Intelligent identification of maceral components of coal based on image segmentation and classification. **Applied Sciences**, 9(16):1-15, 2019.
- WERE, K. et al. A comparative assessment of support vector regression artificial neural networks and random forests for predicting and mapping soil organic carbon stocks across an Afrotropical landscape. **Ecological Indicators**, 52:394-403, 2015.
- WU, W.; DANDY, G. C.; MAIER, H. R. Protocol for developing ANN models and its application to the assessment of the quality of the ANN model development process in drinking water quality modelling. **Environmental Modelling & Software**, 54:108-127, 2014.
- XIE, Z. et al. Artificial intelligence for rapid identification of the coronavirus disease 2019 (COVID-19). **medRxiv**, e20062661, 2020.
- ZHOU, X. et al. Estimation of biomass in wheat using random forest regression algorithm and remote sensing data. **The Crop Journal**, 4(3):212-219, 2016.