Prediction of Soil colour using vis-NIR spectroscopy and machine learning models

Abstract:

Soil color is a key indicator of soil properties and conditions, exerting influence on both agronomic andenvironmental variables. Global Positioning System(GPS) based 2216 surface soil (0-15cm) samples were collected from Kymore Plateau and Satpura hill zone of Madhya Pradesh. The soil colour parameters were recorded using Munsell soil colour chart in the field and soil chemical analysis using standard in laboratory procedures. Spectra of soil samples were alsorecorded using Spectro-radiometer in the laboratory condition. The results obtained in study area indicated that the soil colour component i.e., Hue varied from 10R, 10YR, 2.5Y, 2.5YR, 5Y, 5R 5YR, and 7.5YR, Values and Chroma varied from 2 to 7 and 1 to 8.

Correlation results indicated that the R, G, and B were negatively correlated with organic carbon r = -0.114**, r=-0.071** and r=-0.101*, respectively. The best fitted with the models were polynomial relationship between the value and Chroma of colour parameter and SOC, equation of $Y=0.086x^2-0.860x+7.528$ with a R^2 is 0.982 and $Y=0.018x^2-0.249x+6.126$ with a R^2 is 0.948, respectively. However, The linear relationship was found between the chroma and available P, Y=-0.873+13.92 with a R^2 is 0.922. The Munsell parameters observed in the field were studied using different machine learning models *viz.*,PLSR, SVM, Random Forest, ANN, XGBoost, LightGBM, CatBoost and ELM algorithms. The results showed the Random Forest and XGBoost algorithms gave the best prediction of soil colour parameters L*, a*, b*, R, G, and B with a model accuracy of 83.6, 80.9, 83.0, 84.3, 83.7 and 83.4 %, respectively.

Key words: Soil colour, Spectroscopic, Machin learning model, Munsell colour chart, GPS

1.0 Introduction

The soil color holds significant importance in the realm of soil science, frequently documented in soil profile descriptions due to its role as an initial indicator of various soil conditions and characteristics. Observations made in field studies reveal that soil color is not only readily discernible but also closely associated with numerous other soil properties. It functions effectively as a rapid assessment tool for parameters including drainage class, soil classification, and organic carbon content, all of which may be influenced by the application of variable rate fertilizers and pesticides, as indicated by Moritsuka et al. (2019) and Liu et al. (2020). Established correlations have been noted between soil color and soil texture, water content, iron content, and organic carbon levels. Generally, darker surface soils correlate with elevated organic matter content, which implies fertile conditions that are conducive to plant growth. Such soils are often perceived to possess favorable qualities, including good drainage, adequate aeration, high nitrogen content, and lower susceptibility to erosion. Conversely, lighter-colored soils are typically associated with contrasting characteristics.

The characterization of soil color is commonly performed utilizing the Munsell color system, a method that has gained widespread acceptance among soil surveyors for classification purposes, thereby serving as a fundamental tool in the field of soil descriptions. Although the Munsell system remains dominant, alternative color models, such as the RGB (Red, Green, Blue) system, are frequently employed in digital contexts. The process of characterizing soil color through Cartesian systems can be conducted directly or achieved through the conversion of Munsell data into these alternative systems via lookup tables or statistical regression, as outlined by Dominguez Soto et al. (2012). Moreover, advancements in technology have significantly improved the efficient and precise acquisition of digital soil color data, facilitating analyses that are more consistent and accurate in comparison to traditional visual methods.

Despite multiple attempts to compile regional maps of soil color, a comprehensive map of topsoil color has yet to materialize, as noted by Poppie et al. (2020). Mapping initiatives have encountered limitations, often relying on interpolations of point-based soil observations from regions in Australia (Viscarra-Rossel et al., 2010), China (Liu et al., 2020), and other locations (Soils, 2023). Due to the discrete nature of soil color, it does not inherently reflect the continuous spatial variability that exists within soils. As a result, interpolation maps generated from point-based data heavily depend on field observations collected at specific times, possibly failing to capture the broader and dynamic aspects of soil color changes, as emphasized by Liu et al. (2020). The absence of fine-scale mapping of surface soil color within the study area underscores the necessity for such efforts to effectively monitor ongoing soil conditions.

2.0 Materials and methods

2.1 Study area: The Kymore plateau and Satpura hills zone selected for study which comprise of Rewa, Satna, Panna, Jabalpur, Seoni, Katni, Sidhi and Singrauli districts covering 49.97 lakh ha (16.26%) area of the state. The region has a relatively high proportion of waste and uncultivated lands-about 21%. Another about 22% of the land is under forest cover. Only 37% is cultivated. Irrigation facilities are very poor as only about ten per cent of the cultivated land is irrigated. Mixed red and medium black soils are mainly in the region. In this region, annual rainfall is in the range of 1100 to 1400 mm and wheat and rice are main crops. GPS based 2216 surface soil samples (0-15 cm) were collected from from farmer's fields during the off season of 2022-2023.

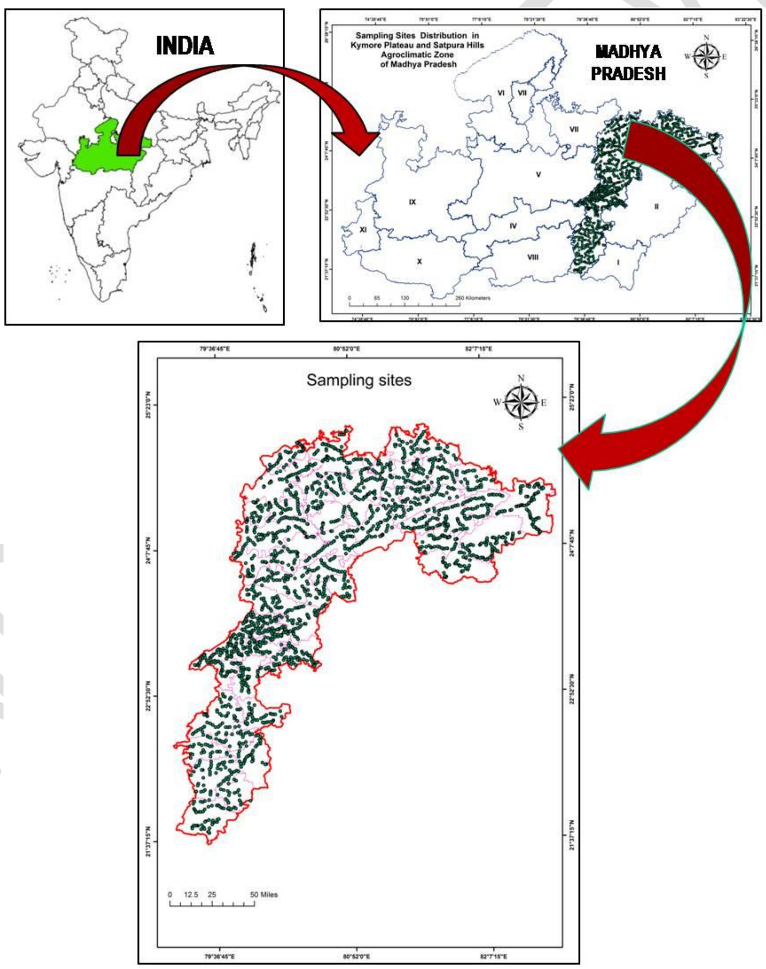


Fig. 1 Location map of study area

2.2 Sample collection and processing

At each sampling point, a 1 kilogram composite soil sample, which accurately represents the area, was collected and recorded in a correctly labeled sample bag. The collected samples were air dried in the shade, crushed using a wooden log to break up clumps and aggregates, and any visible root fragments were removed. Each sample was then passed through a 2 mm sieve and stored in polythene bag with proper labeling for further analysis.

2.3 Soil Colour

The colour parameters of soil (i.e., Hue, Value and Chroma) were recorded by Munsell colour chart (1994 Revised Edition). By placing the sample directly behind the colour sheets separating the nearest matching colour chips, the colour of the soilsamples was matched. Hue, Value, and Chroma—the resulting colour notations—were recorded in the register in relation to the sample code.Conversion- HVC to L* b* a* R G B colour parameter as described by Kirillova *et al*., (2018).

2.4 Spectral reflectance measurements

The spectra reflectance of soil samples recorded with the help of Spectroradiometer (RS-3500). Air-dried samples were placed evenly on a rectangular black disk of 5 cm diameter and 2 cm depth by tapping the tray on a table to ensure a smooth surface. The soil-filled rectangular black disk was kept on the dark background of a table and light reflectance was measured. Reflectance measurements were taken under dark room conditions and colour estimates were made in the 350 to 2500 nm wavelength region of the spectrum.

2.5 Model building

The Hyperspectral data were analyzed to determine RGB and correlated with Munsell Soil colour measurement using correlation techniques. The PLSR, SVR, ANN, Random Forest, XGBoost, Light GBM, Cat Boost machine learning algorithms to develop the model for the prediction of soil colour parameters, the whole data set (n=2217) was divided 1:5 ratios (every fifth sample used for testing dataset) into two datasets viz., training dataset (n=1774) and testing dataset (n=443) in the 80:20 ratios. The training dataset was used to derive the spectral model and the testing dataset was applied to check the predictive performance of the newly developed model.Evaluate the model's performance using testing dataset metrics such as R^2, RMSE and RPD values.

3.0 Result and discussion

3.1 Variation of soil colour

The Munsell soil colour parameters which were presented in Table 1 and depicted in Fig 2,3 and 4 showed that the Hue 10R, 10YR, 2.5Y, 2.5YR, 5Y, 5R 5YR, and 7.5YR and the values and chroma ranged from 2 to 7and 1 to 8 indicating a high colour variation of soil samples collected for the study area. soil colour parameter Hue, about 0, 96, 241, 12, 0, 19, 3 and 8; 0, 86, 75, 4, 0, 0, and 48; 0, 82, 132, 22, 0, 10, 9 and 46; 0, 46, 200, 12, 0, 0, 6 and 20; 0, 71, 151, 12, 3, 4, 8 and 51; 4, 89, 131, 28, 0, 0, 10 and 38; 1, 61, 84, 16, 0, 0, 5 and 43 and 0, 80, 60, 7, 0, 0, 18, and 57 number of samples were observer 10R, 10YR, 2.5Y, 2.5YR, 5R, 5Y, 5YR and 7.5YR, respectively. About 13, 34, 99, 122, 77, 30 and 4; 2, 6, 63, 71, 59, 19 and 0; 0, 1, 34, 101, 116, 43 and 6; 0, 3, 32, 96, 126, 27 and 0; 1, 2, 24, 81, 128, 58 and 6; 13, 22, 130, 88, 41, 6, 0; 1, 2, 31, 67, 75, 30 and 4 and 0, 1, 35, 112, 69, 5 and 0 number of samples fall under value 2, 2.5, 3, 4, 5, 6 and 7, respectively. However, 106, 95, 59, 92, 26 and 1; 11, 32, 43, 93, 41 and 0; 10, 30, 69, 105, 79 and 8; 18, 59, 51, 107, 49 and 0; 16, 39, 58, 117, 69 and 1; 36, 69, 82, 87, 26 and 0; 6, 18, 31, 91, 63 and 1 and 1, 14, 51, 108, 47 and 1 number of samples fall under chroma 1, 2, 3, 4, 6 and 8, respectively in the soils of Jabalpur, Ktani, Panna, Rewa, Satna Seoni, Sidhi and Singrauli Districts. The variation in soil colour brought to be differences in texture, topography, mineralogical and chemical composition and soil moisture regimes (Dinesh *et al.* (2017) and Arya *et al.* (2024).

Table: 1 Spatial distribution of soil colour in Kymore plateau and Satpura hills zone

Kymore plateau and Satpura hills	**n**	**Hue**								**Value**	**Chroma**
		10R	**10YR**	**2.5Y**	**2.5YR**	**5R**	**5Y**	**5YR**	**7.5YR**		
Jabalpur	379	0	96	241	12	0	19	3	8	2-7	1-8
Katni	220	0	86	75	4	0	0	7	48	2-7	2-6
Panna	301	0	82	132	22	0	10	9	46	2.5-7	1-8
Rewa	284	0	46	200	12	0	0	6	20	2.5-6	1-6
Satna	300	0	71	151	12	3	4	8	51	2-7	1-8
Seoni	300	4	89	131	28	0	0	10	38	2-6	2-6
Sidhi	210	1	61	84	16	0	0	5	43	2-6	1-8
Singrauli	222	0	80	60	7	0	0	18	57	2.5-6	1-8
Total	2216	5	611	1074	113	3	33	66	311	2-7	1-8

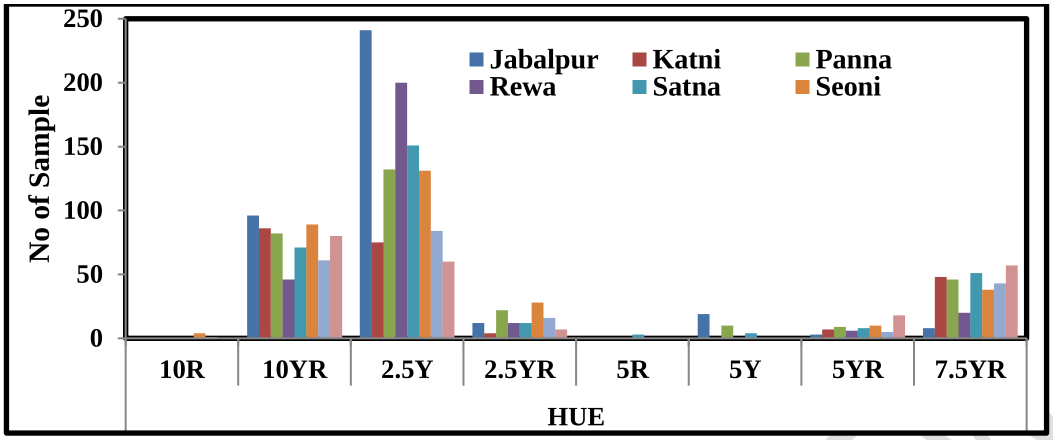


Fig. 2 Frequency distribution of Hue in Kymore plateau and Satpura hills zone

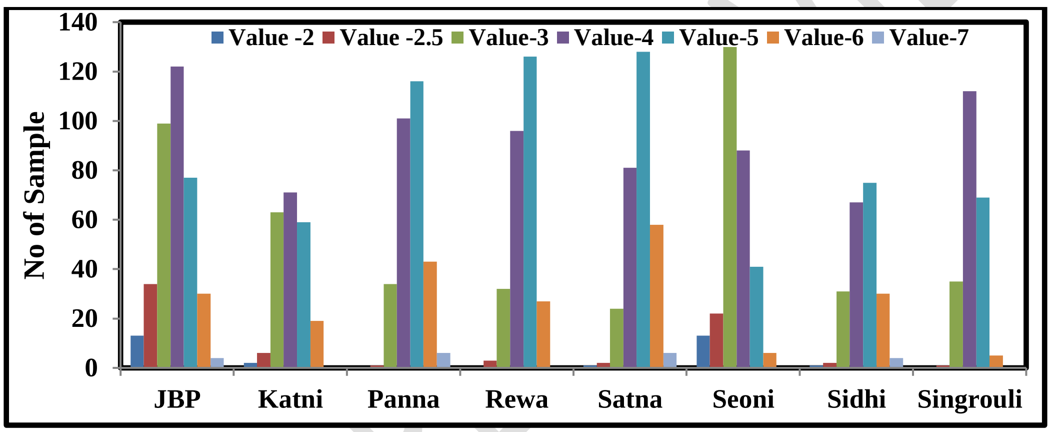


Fig. 3 Frequency distribution of Value in Kymore plateau and Satpura hills zone

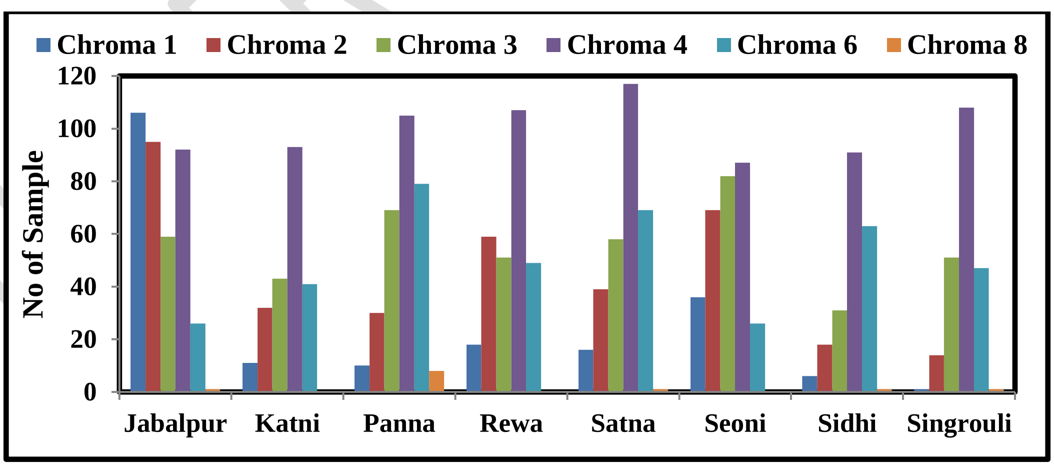


Fig. 4 Frequency distribution of Chroma in Kymore plateau and Satpura hills zone

3.2 Status of physico-chemical properties under different colour parameters

3.2.1pH, EC and SOC content in soils under different Hue

Data presented in Table 2 showed that the pH ranged from 5.82 to 7.30, 4.86 to 8.17, 4.79 to 8.22, 5.00 to 8.02, 6.30 to 7.17, 5.56 to 8.10, 5.24 to 7.92 and 6.55 to 8.39 with a mean value of 6.57, 6.64, 6.82, 6.53, 6.78, 7.02, 6.53 and 6.61at hue of 10R, 10YR, 2.5Y, 2.5YR, 5R, 5Y, 5YR and 7.5YR, respectively. The EC ranged from 0.14 to 0.29, 0.05 to 0.96,0.03 to 0.97, 0.05 to 0.54, 0.19 to .22, 0.08 to 0.66, 0.08 to 0.73 and 0.04 to 0.97 dS/m with a mean value 0.20, 0.25, 0.26, 0.21, 0.20, 0.25, 0.24 and 0.26 dS/m at hue of 10R, 10YR, 2.5Y, 2.5YR, 5R, 5Y, 5YR and 7.5YR, respectively. However, the SOC ranged from 4.70 to 6.98, 1.50 to 12.45,1.50 to 10.95, 1.20 to 9.75,4.73 to 6.03, 1.50 to 7.05, 2.10 to 11.70 and 1.05 to 10.65 g/kg with a mean value of 5.56, 5.48, 5.64, 5.27, 5.55, 4.77, 5.43 and 5.35 g/kg at hue of 10R, 10YR, 2.5Y, 2.5YR, 5R, 5Y, 5YR and 7.5YR, respectively.

3.2.2 pH, EC and SOC content in soils under different Value

The result on pH EC and SOC under different value was presented in Table 2 showed that The pH ranged from5.95 to 7.81, 5.25 to 8.10, 5.00 to 8.10, 4.55 to 8.39, 4.79 to 8.22, 4.87 to 8.11 and 5.41 with a mean value of 6.94, 6.96, 6.70, 6.68, 6.72, 6.75 and 6.91 at value of 2, 2.5, 3, 4, 5, 6 and 7, respectively. The EC ranged from0.09 to 0.92, 0.09 to 0.58, 0.05 to 0.85, 0.03 to 0.97, 0.05 to 0.97, 0.05 to 0.91 and 0.07 to 0.53 dS/m with a mean value of 0.24, 0.26, 0.24, 0.25, 0.26, 0.25 and 0.26dS/m at value of 2, 2.5, 3, 4, 5, 6 and 7, respectively. The SOC ranged from 2.70 to 11.25, 1.95 to 10.65,1.50 to 12.45, 1.05 to 10.9, 1.05 to 10.95, 2.25 to 10.35 and 4.42 to 9.30 g/kg with a mean value of 6.20, 5.89,5.66, 4.49, 5.40 5.47 and 5.71 g/kg at value of 2, 2.5, 3, 4, 5, 6 and 7, respectively.

3.2.3 pH, EC and SOC content in soils under different Chroma

Further from the Table 2 showed that the pH ranged from4.95 to 8.10, 4.79 to 8.22, 4.55 to 8.15, 4.94 to 8.12 and 5.79 to 7.695.14 to 8.39 with a mean value of 7.00, 7.85, 6.73, 6.62, 6.63 and 6.93 at chroma of 1, 2, 3, 4, 6 and 8, respectively. The EC ranged from 0.03 to 0.92, 0.05 to 0.73, 0.07 to 0.82, 0.04 to 0.97, 0.05 to 0.96 and 0.09 to 0.43 dS/m with a mean value of 0.25, 0.26, 0.25,0.25, 0.25 and 0.28dS/m at chroma of 1, 2, 3, 4, 6 and 8, respectively. The OC ranged from 1.50 to 12.45, 1.50 to 10.95, 1.65 to 10.80, 1.05 to 11.70, 1.05 to 9.30 and 3.60 to 7.95g/kg with a mean value of 5.84, 5.79, 5.53,5.44, 5.26 and 5.36 g/kg at chroma of 1, 2, 3, 4, 6 and 8, respectively.

Table 2: Physico-chemical parameter of soils under different Munsell colour parameters

Parameter		n	pH				EC(dSm^{-1})				OC (g kg^{-1})			
			Min	**Max**	**Mean**	**CV(%)**	**Min**	**Max**	**Mean**	**CV(%)**	**Min**	**Max**	**Mean**	**CV(%)**
Hue	**10R**	5	5.82	7.30	6.57	8.83	0.14	0.29	0.20	35.46	4.70	6.98	5.56	15.73
	10YR	611	4.86	8.17	6.64	10.27	0.05	0.96	0.25	49.52	1.50	12.45	5.48	23.49
	2.5Y	1074	4.79	8.22	6.82	9.20	0.03	0.97	0.26	49.82	1.50	10.95	5.64	22.86
	2.5YR	113	5.00	8.02	6.53	9.61	0.05	0.54	0.21	45.68	1.20	9.75	5.27	23.01
	5R	3	6.30	7.17	6.78	6.52	0.19	0.22	0.20	7.51	4.73	6.03	5.55	12.86
	5Y	33	5.36	8.10	7.02	8.91	0.08	0.66	0.25	51.76	1.50	7.05	4.77	24.10
	5YR	66	5.24	7.92	6.53	9.49	0.08	0.73	0.24	58.25	2.10	11.70	5.43	28.70
	7.5YR	311	4.55	8.39	6.61	9.95	0.04	0.97	0.26	58.07	1.05	10.65	5.35	20.93
Value	**2**	30	5.95	7.81	6.94	7.13	0.09	0.92	0.24	74.12	2.70	11.25	6.20	26.18
	2.5	71	5.25	8.10	6.96	8.70	0.09	0.58	0.26	41.37	1.95	10.65	5.89	26.94
	3	448	5.00	8.10	6.70	9.35	0.05	0.85	0.24	49.99	1.50	12.45	5.66	23.67
	4	738	4.55	8.39	6.68	9.79	0.03	0.97	0.25	51.72	1.05	10.95	5.49	23.16
	5	691	4.79	8.22	6.72	9.68	0.05	0.97	0.26	51.15	1.05	10.95	5.40	22.80
	6	218	4.87	8.11	6.75	11.02	0.05	0.91	0.25	52.99	2.25	10.35	5.47	19.10
	7	20	5.41	7.82	6.91	9.06	0.07	0.53	0.26	48.48	4.42	9.30	5.71	22.85
Chroma	**1**	204	4.95	8.10	7.00	8.37	0.03	0.92	0.25	50.86	1.50	12.45	5.84	27.58
	2	356	4.79	8.22	6.85	9.36	0.05	0.73	0.26	47.20	1.50	10.95	5.79	24.34
	3	444	5.14	8.39	6.73	9.61	0.07	0.82	0.25	47.96	1.65	10.80	5.53	22.43
	4	800	4.55	8.15	6.62	9.89	0.04	0.97	0.25	54.75	1.05	11.70	5.44	21.98
	6	400	4.94	8.12	6.63	9.87	0.05	0.96	0.25	52.49	1.05	9.30	5.26	20.41
	8	12	5.79	7.69	6.93	8.35	0.09	0.43	0.28	36.17	3.60	7.95	5.36	19.15

3.3 Status of macro nutrients in soils under different colour parameters

3.3.1 Availability of P, K and S in soils under different Hue

Data given in Table 3 revealed that the available P ranged from 13.59 to 40.79, 1.44 to 74.02, 1.15 to 117.17, 1.19 to 74.02, 5.17 to 7.77, 3.43 to 29.22, 1.73 to 26.32 and 1.11 to 67.12 kg ha^{-1}with a mean value of 20.65, 10.22, 11.20, 10.59, 6.32, 9.97, 8.74 and 9.86 kg ha^{-1}at hue of 10R, 10YR, 2.5Y, 2.5YR, 5R, 5Y, 5YR and 7.5YR, respectively. The available K ranged from 290.08 to 386.40, 116.48 to 715.68, 105.28 to 974.40, 122.08 to 611.52,244.16 to 324.80, 126.56 to 557.76, 147.84 to 611.52 and 145.60 to 750.40 kg ha^{-1}with a mean value of 343, 344, 366, 340, 293, 355, 324 and 327 kg ha^{-1}at hue of 10R, 10YR, 2.5Y, 2.5YR, 5R, 5Y, 5YR and 7.5YR, respectively. The available S ranged from 3.12 to 8.79, 0.43 to 33.60, 0.57 to 30.77, 0.57 to 19.00, 5.10 to 7.66, 2.55 to 19.99, 0.99 to 20.84 and 0.71 to 18.29 mgkg^{-1}with a mean value of 7.34, 7.71, 8.21, 6.95, 6.57, 9.04, 6.82 and 6.78 mg kg^{-1}at hue of 10R, 10YR, 2.5Y, 2.5YR, 5R, 5Y, 5YR and 7.5YR, respectively.

3.3.2 Availability of P, K and S in soils under different Value

From the result presented in Table 3 showed the available P ranged from 1.73 to 50.05, 1.44 to 67.12, 1.44 to 61.62, 1.19 to 96.30, 1.11 to 117.7, 1.44 to 74.02 and 2.89 to 28.64 kg ha^{-1} with a mean value of 15.32, 13.56, 10.35, 10.23, 10.61, 11.05 and 10.83 kg ha^{-1} at value of 2, 2.5, 3, 4, 5, 6 and 7, respectively. The available K ranged from 136.64 to 624.96, 178.08 to 750.40,116.48 to 678.72, 125.44 to 738.08, 122.08 to 848.96, 160.16 to 974.40 and 105.28 to 468.16 kg ha^{-1} with a mean value of376, 388, 355, 349, 348, 354 and 301 kg ha^{-1} at value of 2, 2.5, 3, 4, 5, 6 and 7, respectively. The available S ranged from 2.13 to 33.60, 2.55 to 30.77, 0.57 to 29.77, 0.43 to 30.34, 0.57 to 26.23, 0.85 to 27.51 and 2.41 to 18.57 mg $kg^{-1}$with a mean value of 8.69, 7.96, 7.66, 7.84, 7.62, 8.03 and 8.02 mg kg^{-1}at value of 2, 2.5, 3, 4, 5, 6 and 7, respectively.

3.3.3Availability of P, K and S in soils under different chroma

Data on available P, K and S in Table 3 indicated that the available P ranged from 1.44 to 96.30, 1.44 to 95.18, 1.11 to 74.02, 1.15 to 55.84, 1.19 to 117.17 and 2.02 to 11.28 kg $ha^{-1}$with a mean value of13.73, 12.32, 10.22, 10.05, 9.30 and 6.94 kg ha^{-1}at chroma of1, 2, 3, 4, 6 and 8, respectively. The available K ranged from 136.64 to 848.96, 126.56 to 750.40, 116.48 to 721.28, 114.24 to 743.68, 105.28to 974.40 and 252.00 to 543.20kg $ha^{-1}$with a mean value of 397, 376, 348, 341, 329 and 389 kg ha^{-1}atchroma of1, 2, 3, 4, 6 and 8, respectively. The available S ranged from 1.13 to 33.60, 0.71 to 30.34,0.85 to 29.77 0.43 to 25.38, 0.57 to 23.25 and 2.98 to 22.12 mg $kg^{-1}$with a mean value of 8.69, 8.51, 7.82,7.54, 7.07 and 7.40 mg kg^{-1}at chroma of 1, 2, 3, 4, 6 and 8, respectively.

Table 3: Macro-nutrients status under different Munsell colour parameters

Parameter		n	P (kg ha^{-1})				K (kg ha^{-1})				S (mg kg^{-1})			
			Min	Max	Mean	CV (%)	Min	Max	Mean	CV (%)	Min	Max	Mean	CV (%)
Hue	10R	5	13.59	40.79	20.65	56.22	290	386	343	11.10	3.12	8.79	7.34	32.49
	10YR	611	1.44	74.02	10.22	75.02	116	716	344	28.73	0.43	33.60	7.71	48.35
	2.5Y	1074	1.15	117.17	11.20	85.06	105	974	366	29.89	0.57	30.77	8.21	48.71
	2.5YR	113	1.19	74.02	10.59	86.10	122	612	340	29.79	0.57	19.00	6.95	45.25
	5R	3	5.17	7.77	6.32	20.97	244	325	293	14.72	5.10	7.66	6.57	20.05
	5Y	33	3.43	29.22	9.97	57.03	127	558	355	29.64	2.55	19.99	9.04	44.90
	5YR	66	1.73	26.32	8.74	63.87	148	612	324	30.17	0.99	20.84	6.82	43.32
	7.5YR	311	1.11	67.12	9.86	74.13	146	750	327	26.77	0.71	18.29	6.78	44.80
Value	2	30	1.73	50.05	15.32	81.44	137	625	376	34.22	2.13	33.60	8.69	71.93
	2.5	71	1.44	67.12	13.56	86.10	178	750	388	30.43	2.55	30.77	7.96	49.82
	3	448	1.44	61.62	10.35	65.95	116	679	355	27.46	0.57	29.77	7.66	45.45
	4	738	1.19	96.30	10.23	81.50	125	738	349	30.41	0.43	30.34	7.84	50.37
	5	691	1.11	117.17	10.61	84.10	122	849	348	28.68	0.57	26.23	7.62	46.02
	6	218	1.44	74.02	11.05	88.82	160	974	354	30.90	0.85	27.51	8.03	48.82
	7	20	2.89	28.64	10.83	59.63	105	468	301	30.61	2.41	18.57	8.02	55.19
Chroma	1	204	1.44	96.30	13.73	91.37	137	849	397	30.13	1.13	33.60	8.69	54.51
	2	356	1.44	95.18	12.32	82.47	127	750	376	29.37	0.71	30.34	8.51	47.97
	3	444	1.11	74.02	10.22	76.45	116	721	348	29.64	0.85	29.77	7.82	46.52
	4	800	1.15	55.84	10.05	69.30	114	744	341	28.43	0.43	25.38	7.54	45.45
	6	400	1.19	117.17	9.30	85.99	105	974	329	27.56	0.57	23.25	7.07	49.14
	8	12	2.02	11.28	6.94	47.45	252	543	389	24.15	2.98	22.12	7.40	67.05

3.4 Correlation findings

Correlation result showed in Table 4 indicated that the CIE lab parameters *i.e.,* L* was significantly correlated with SOC (r=-0.088**) and K (r= -0.062**). Further, a* and b* parameters were found significantly correlated with all the parameters except EC which was showed non-significant correlation with b*. The SOC was significantly related with R, G and B with r value of -0.114**, -0.071**) and-0.051*) but EC showednon-significant correlation with R, G and B colour component, respectively. Munsell parameters i.e., Value showed significant relation with OC (r=-0.82**) and available K (r=-0.51*). However, chroma was negatively correlated with OC (r=-0.141**), available P (r=-0.143**), K (r=-0.173**) and S (r=-0.134**). Minh *et al.,* (2023), Djama *et al.,* (2023) and Kang*et al.,* (2024) also found negatively significant correlation between SOC and soil color component (RGB).

Table 4: Correlation between soil colour and soil chemical properties

Parameters	**Munsell parameters**		**CIE-Lab color space**			**RGB parameters**		
	Value	**Chroma**	**L***	**a***	**b***	**Red**	**Green**	**Blue**
pH	-0.002^{NS}	-0.149**	-0.015^{NS}	-0.160**	-0.120**	-0.065**	0.004^{NS}	0.056**
EC	0.028^{NS}	-0.013^{NS}	0.019^{NS}	-0.058**	-0.016^{NS}	0.002^{NS}	0.029^{NS}	0.017^{NS}
OC	-0.082**	-0.141**	-0.088**	-0.115**	-0.105**	-0.114**	-0.071**	-0.101**
Available P	-0.015^{NS}	-0.143**	-0.019^{NS}	-0.095**	-0.106**	-0.053*	-0.002^{NS}	0.055*
Available K	-0.051*	-0.173**	-0.062**	-0.148**	-0.117**	-0.098**	-0.045*	-0.003^{NS}
Available S	-0.000^{NS}	-0.134**	-0.001^{NS}	-0.158**	-0.089**	-0.042^{NS}	0.017^{NS}	0.048*

*= significant at 5%, **= significant at 1%,   NS = Non significant

Further, the Fig. 2, 3 and 4 depicted the relationship between soil color parameter value and chroma. The best fitted with the models were polynomial relationship between the value and Chroma of colour parameter and SOC, equation of $Y=0.086x^2-0.860x+7.528$ with a R^2 is 0.982 and $Y=0.018x^2-0.249x+6.126$ with a R^2 is 0.948, respectively. However, The linear relationship was found between the chroma and available P, $Y=-0.873+13.92$ with a R^2 is 0.922.

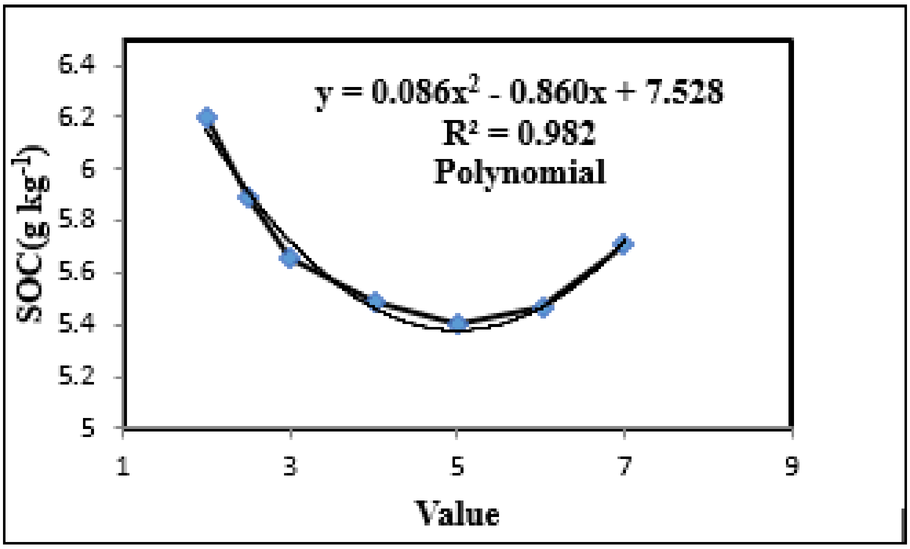


Fig.2 Relationship between Value and SOC

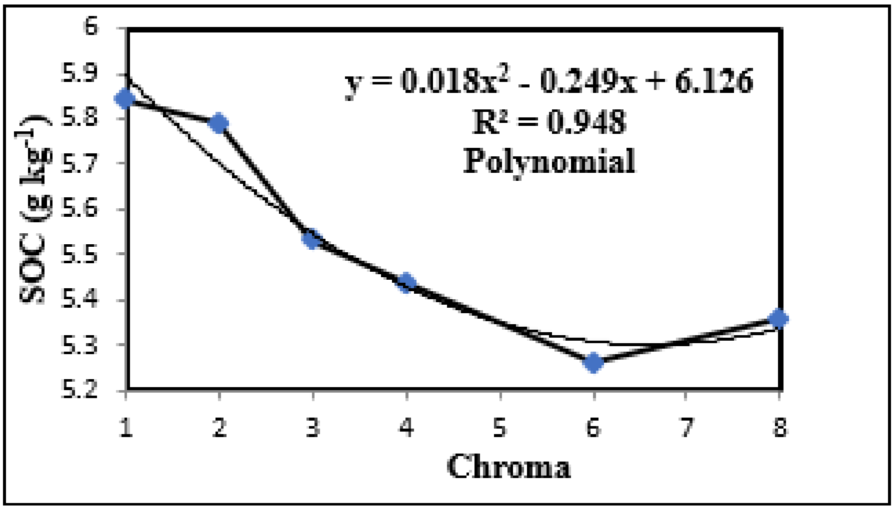


Fig. 3 Relationship between Chroma and SOC

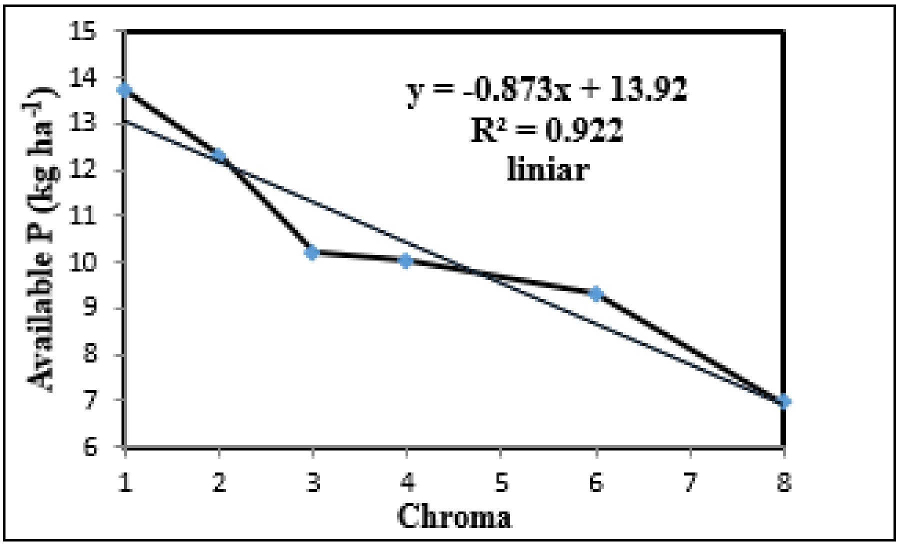


Fig.4 Relationship between Chroma and available P

3.5 Prediction of soil colour parameters using different machine learning models

3.5.1 CIE-Lab color space

The performance results presented in the Table 5 showed the PLSR, SVR, ANN, Random Forest, XGBoost, LightGBM, CatBoost and ELM models showed the R^2 of 0.24, 0.18, 0.28, 0.84, 0.84, 0.84, 0.83 and 0.38; 0.32, 0.28, 0.38, 0.81, 0.81, 0.81, 0.81 and 0.31 and 0.20, 0.12, 0.23, 0.83, 0.83, 0.83, 0.82 and 0.20 with a RMSE and RPD of 9.43, 9.79, 9.14, 5.25, 5.28, 5.26, 5.30 and 9.46; 3.99, 4.40, 3.83, 2.37, 2.37, 2.38, 2.35 and 4.04; 8.48, 9.03, 8.32, 4.85, 4.82, 4.84, 4.86 and 8.50 and 1.14, 1.10, 1.18, 2.06, 2.05, 2.05, 2.04 and 1.14; 1.22, 1.10, 1.27, 2.04, 2.05, 2.04, 2.07 and 1.20 and 1.12, 1.05, 1.14, 1.96, 1.97, 1.96, 1.95 and 1.12 for L*, a* and b* for a training dataset. However, PLSR, SVR, ANN, Random Forest, XGBoost, LightGBM, CatBoost and ELM models showed R^2 of 0.33, 0.29, 0.36, 0.36, 0.36, 0.35, 0.36 and 0.35; 0.37, 0.33, 0.35, 0.31, 0.31, 0.32, 0.32 and 0.32 and 0.30, 0.22, 0.24, 0.36, 0.37, 0.37, 0.36 and 0.31 with a RMSE and RPD of 7.23, 7.41, 7.11, 7.10, 7.07, 7.17, 7.11 and 7.14; 3.17, 3.41, 3.27, 3.33, 3.33, 3.29, 3.29 and 3.30 and 6.36, 6.72, 6.72, 6.05, 6.03, 6.06, 6.08 and 6.31 and 1.22, 1.19, 1.24, 1.24, 1.25, 1.23, 1.24and1.23; 1.25, 1.16, 1.21, 1.21, 1.19, 1.20, 1.20 and 1.20 and 1.19, 1.12, 1.12, 1.25, 1.25, 1.25, 1.24 and 1.20 for the testing dataset, respectively. Overall the XGBoost, Random Forest and XGBoost model gave the best prediction of soil colour parameter of L*, a* and b*.

Table 5: Results of different model for prediction of soil colour parameter (L*, a* and b*)

Property	Model	Training (80%)			Testing data (20%)		
		R^2	RMSE	RPD	R^2	RMSE	RPD
L*	PLSR	0.24	9.43	1.14	0.33	7.23	1.22
	SVR	0.18	9.79	1.10	0.29	7.41	1.19
	ANN	0.28	9.14	1.18	0.36	7.11	1.24
	RandomForest	0.84	5.25	2.06	0.36	7.10	1.24
	XGBoost	**0.84**	**5.28**	**2.05**	**0.36**	**7.07**	**1.25**
	LightGBM	0.84	5.26	2.05	0.35	7.17	1.23
	CatBoost	0.83	5.30	2.04	0.36	7.11	1.24
	ELM	0.23	9.46	1.14	0.35	7.14	1.23
a*	PLSR	0.32	3.99	1.22	0.37	3.17	1.25
	SVR	0.28	4.40	1.10	0.33	3.41	1.16
	ANN	0.38	3.83	1.27	0.35	3.27	1.21
	RandomForest	**0.81**	**2.37**	**2.04**	**0.31**	**3.33**	**1.21**
	XGBoost	0.81	2.37	2.05	0.31	3.33	1.19
	LightGBM	0.81	2.38	2.04	0.32	3.29	1.20
	CatBoost	0.81	2.35	2.07	0.32	3.29	1.20
	ELM	0.31	4.04	1.20	0.32	3.30	1.20
b*	PLSR	0.20	8.48	1.12	0.30	6.36	1.19
	SVR	0.12	9.03	1.05	0.22	6.72	1.12
	ANN	0.23	8.32	1.14	0.24	6.72	1.12
	RandomForest	0.83	4.85	1.96	0.36	6.05	1.25
	XGBoost	**0.83**	**4.82**	**1.97**	**0.37**	**6.03**	**1.25**
	LightGBM	0.83	4.84	1.96	0.37	6.06	1.25
	CatBoost	0.82	4.86	1.95	0.36	6.08	1.24
	ELM	0.20	8.50	1.12	0.31	6.31	1.20

3.5.2 Spectral colour-R, G, B

The performance results presented in the Table 6 indicated that the PLSR, SVR, ANN, Random Forest, XGBoost, LightGBM, CatBoost and ELM models showed the R^2 of 0.25, 0.18, 0.37, 0.84, 0.85, 0.84, 0.84 and 0.0.25; 0.23, 0.16, 0.29, 0.84, 0.84, 0.84, 0.84 and 0.22 and 0.16, 0.08, 0.16, 0.84, 0.83, 0.83, 0.84 and 0.15 with a RMSE and RPD of 25.60, 26.77, 23.60, 14.05, 14.01, 14.12, 14.04 and 25.67; 22.18, 23.22, 21.43, 12.41, 12.32, 12.38, 12.40 and 22.30; 19.96, 20.93, 20.08, 11.38, 11.37, 11.38, 11.34 and 20.12 and 1.15, 1.10, 1.25, 2.10, 2.11, 2.09, 2.11 and 1.15; 1.14, 1.09, 1.18, 2.04, 2.06, 2.05, 2.04 and 1.14 and 1.09, 1.04, 1.09, 1.92, 1.92, 1.92, 1.92 and 1.08 for R, G and B for a training dataset. However, PLSR, SVR, ANN, Random Forest, XGBoost, LightGBM, CatBoost and ELM models showed R^2 of 0.37, 0.37, 0.31, 0.45, 0.45, 0.45, 0.45 and 0.38; 0.39, 0.34, 0.41, 0.43, 0.43, 0.43, 0.43 and 0.42 and 0.23, 0.16, 0.09, 0.30, 0.31, 0.29, 0.30 and 0.22 with a RMSE and RPD of 17.89, 17.83, 19.96, 16.74, 16.68, 16.72, 16.78 and 17.69; 15.23, 15.68, 15.25, 14.70, 14.71, 14.74, 14.68 and 14.85 and 13.61, 14.06, 16.06, 12.98, 12.88, 13.06, 13.03 and 13.70 and 1.25, 1.26, 1.12, 1.34, 1.35, 1.34, 1.34 and1.27; 1.27, 1.24, 1.27, 1.32, 1.32, 1.31, 1.32 and 1.30 and 1.13, 1.09, 0.95, 1.18, 1.19, 1.17, 1.18 and 1.12 for the testing dataset, respectively. Overall the Random Forest, Random Forest and XGBoost model gave the best prediction of soil colour parameter of R, G and B. Pegalajaret al. (2020), Al-Najiet al. (2022), Ramos et al. (2020), Silvero et al. (2021) and de Souzaet al. (2022) also used machine learning algorithms for rapid assessment and prediction of soil color parameters and found some similar trends and supporting finding.

Table 6: Results of different model for prediction of soil colour parameter (R, G and B)

Property	Model	Training data(80%)			Testing data (20%)		
		R^2	RMSE	RPD	R^2	RMSE	RPD
R	PLSR	0.25	25.60	1.15	0.37	17.89	1.25
	SVR	0.18	26.77	1.10	0.37	17.83	1.26
	ANN	0.37	23.60	1.25	0.31	19.96	1.12
	RandomForest	**0.84**	**14.05**	**2.10**	**0.45**	**16.74**	**1.34**
	XGBoost	0.85	14.01	2.11	0.45	16.68	1.35
	LightGBM	0.84	14.12	2.09	0.45	16.72	1.34
	CatBoost	0.84	14.04	2.11	0.45	16.78	1.34
	ELM	0.25	25.67	1.15	0.38	17.69	1.27
G	PLSR	0.23	22.18	1.14	0.39	15.23	1.27
	SVR	0.16	23.22	1.09	0.34	15.68	1.24
	ANN	0.29	21.43	1.18	0.41	15.25	1.27
	RandomForest	**0.84**	**12.41**	**2.04**	**0.43**	**14.70**	**1.32**
	XGBoost	0.84	12.32	2.06	0.43	14.71	1.32
	LightGBM	0.84	12.38	2.05	0.43	14.74	1.31
	CatBoost	0.84	12.40	2.04	0.43	14.68	1.32
	ELM	0.22	22.30	1.14	0.42	14.85	1.30
B	PLSR	0.16	19.96	1.09	0.23	13.61	1.13
	SVR	0.08	20.93	1.04	0.16	14.06	1.09
	ANN	0.16	20.08	1.09	0.09	16.06	0.95
	RandomForest	0.84	11.38	1.92	0.30	12.98	1.18
	XGBoost	**0.83**	**11.37**	**1.92**	**0.31**	**12.88**	**1.19**
	LightGBM	0.83	11.38	1.92	0.29	13.06	1.17
	CatBoost	0.84	11.34	1.92	0.30	13.03	1.18
	ELM	0.15	20.12	1.08	0.22	13.70	1.12

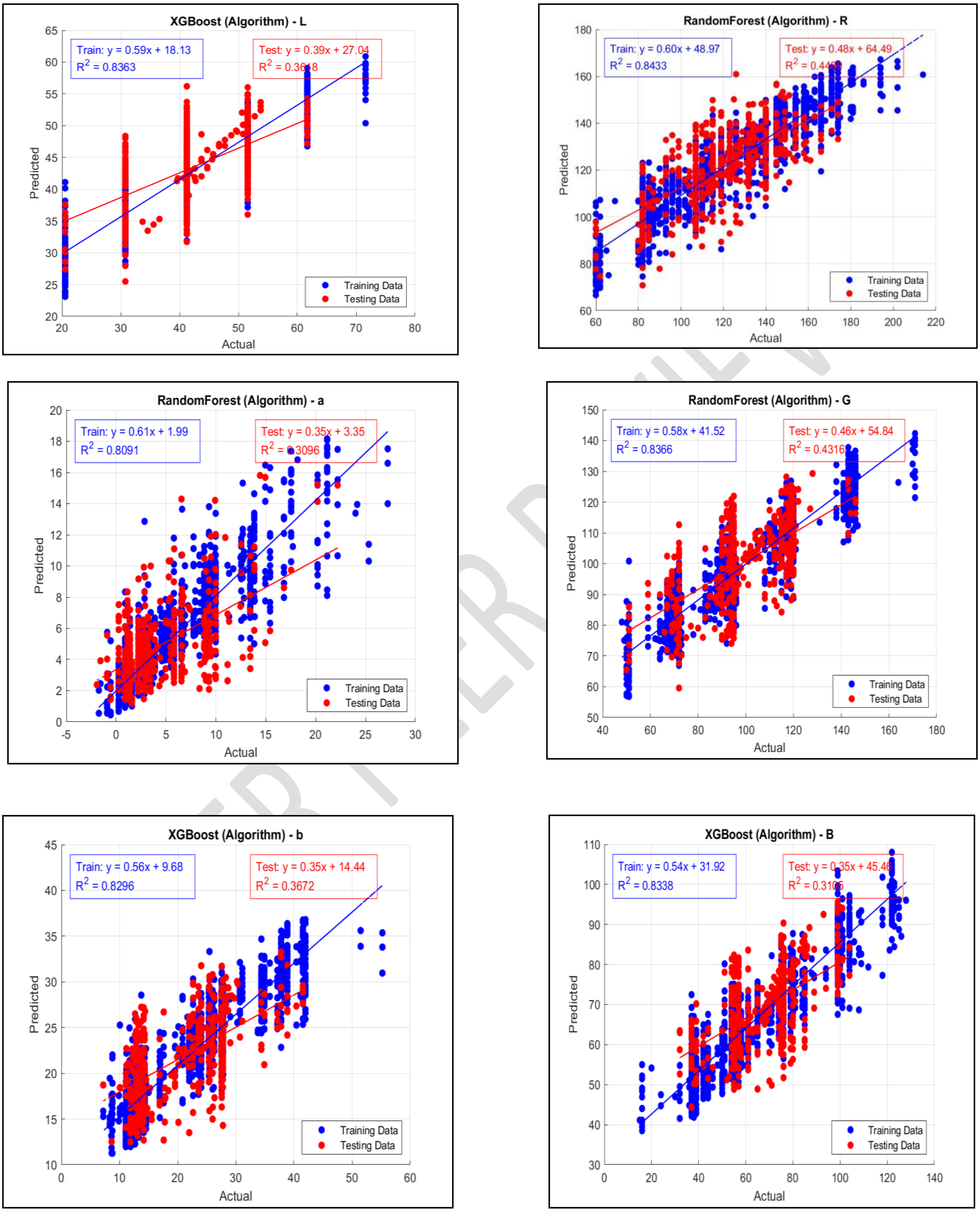
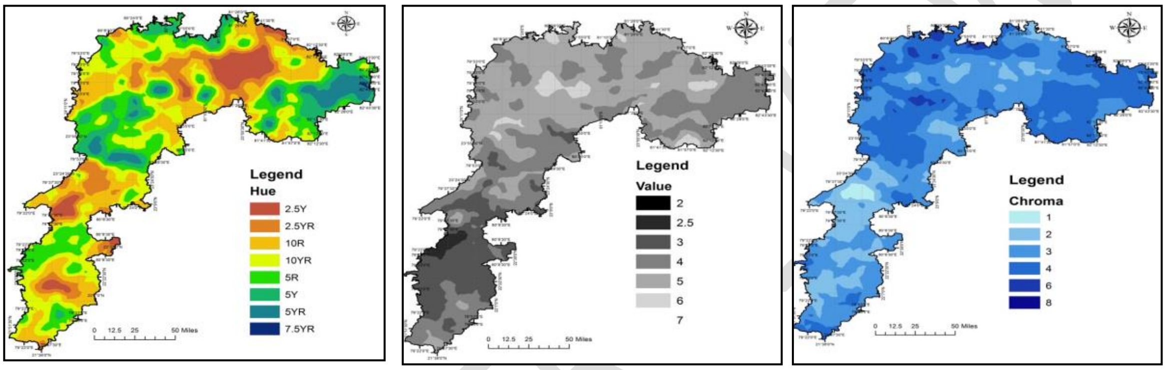


Fig. 5 Prediction performance of color parameters (L*, a*, b*, R, G and B)

3.6 GIS based mapping of soil colour parameters

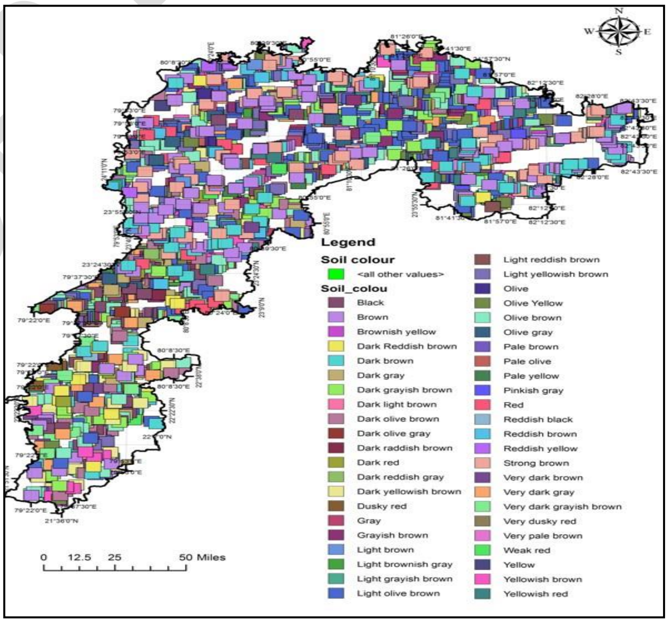
Spatial variability map for hue, value and chroma (Munsell colour parameters) are shown in Fig. 6. The hue 2.5Y was highly distributed in the north region than the south and 5YR and 7.5YR varied in the southern part. The higher the value in north region and lower value was observed in the southern part. Low values are usually due to the presence of organic matter or Mn oxides Rizzo et al., (2023). Chroma was lower distributed in south - west in the soils of Kymore plateau and Satpura hill zone of Madhya Pradesh.

Fig. 6 Spatial variability map of soil colour parameter



(a) Hue **(b) Value** **(c) Chroma**

(d) Soil colour types



4.0 Conclusions

Variation in soil colour was recorded using hyper-spectral data and machine learning models. The RGB values showed apparent differences between samples. Correlation study showed the L*, Munsell value and chroma were negatively related with SOC and K. The SOC was negatively correlated with R, G and B. The machine learning results suggested that the Random Forest and XGBoost models were found most potential and gave the best prediction of soil colour parameters. GIS based soil colour maps generated can also serve as covariates for mapping, offering comprehensive insights into the soil's properties.

Reference

4 Al-Naji, A., Fakhri, A. B., Gharghan, S. K., &Chahl, J. (2021). Soil color analysis based on a RGB camera and an artificial neural network towards smart irrigation: A pilot study. Heliyon, 7(1).

5 Arya, S., Singh, R., Jat, M. K., & Yadav, P. K. (2024). Characterization and Classification of Ber (*Ziziphusmauritiana*) Growing Soils of Rewari District, Haryana: Characterization and classification of ber growing soils. Journal of Soil Salinity and Water Quality, 16(1), 1-15.

6 de Souza, W. M., Ribeiro, A. J. A., & da Silva, C. A. U. (2022). Use of ANN and visual-manual classification for prediction of soil properties for paving purposes. International Journal of Pavement Engineering, 23(5), 1482-1490.

7 Dinesh, Bhat, M.A. and Grewal K.S. (2017) Characterization and classification of soils on different geomorphic units of North–Eastern Haryana, India. Agropedology, 27(02): 103–116.

8 Djama, Z. A., Kavaklıgil, S. S., &Erşahin, S. (2023). Evaluation of Soil Color and Soil Fertility Relations on Cultivated Semi-Arid Sloping Landscapes. Journal of Agricultural Faculty of Gaziosmanpaşa University (JAFAG), 40(1), 19-25.

9 Domínguez Soto, J. M., Román Gutiérrez, A. D., Prieto García, F., & Acevedo Sandoval, O. (2012). Munsell notation system and cielab as a tool for evaluation colors in soils. *Revistamexicana de cienciasagrícolas*, *3*(1), 141-155.

10 Kang, Y. G., Lee, J. Y., Kim, J. H., & Oh, T. K. (2024). Quantifying soil organic matter for sustainable agricultural land management with soil color and machine learning technique. Agronomy Journal.

11 Liu, F., Rossiter, D.G., Zhang, G. L. and Li, D.C., (2020). A soil colour map of China. Geoderma 379, 114556.

12 Minh, V. Q., Vu, P. T., Du, T. T., &Tinh, T. K. (2023). The Soil Color as Indicator for Fruit Garden Soil Assessment and Recommendation. Agricultural Engineering International: CIGR Journal, 25(2).

13 Moritsuka, N., Matsuoka, K., Katsura, K., & Yanai, J. (2019). Farm-scale variations in soil color as influenced by organic matter and iron oxides in Japanese paddy fields. Soil Science and Plant Nutrition, *65*(2), 166-175.

14 Moritsuka, N., Matsuoka, K., Katsura, K., Sano, S., &Yanai, J. (2014). Soil color analysis for statistically estimating total carbon, total nitrogen and active iron contents in Japanese agricultural soils. Soil science and plant nutrition, *60*(4), 475-485.

15 N.P. Kirillova (2018) new perspectives to use Munsell color charts with electronic devices Comput. Electron. Agric.

16 Pachlaniya, N. K., Jain, R. K., & Shukla, A. K. (2023). Evaluations of soil fertility status of available major nutrients and micro nutrients in vertisol of indore district of Madhya Pradesh India. Plant archives, 23(1), 100-104.

17 Pegalajar, M. C., Ruíz, L. G. B., Sánchez-Maranon, M., &Mansilla, L. (2020). A Munsellcolour-based approach for soil classification using Fuzzy Logic and Artificial Neural Networks. *FUZZY sets and systems*, *401*, 38-54.

18 Poppiel, R.R., Lacerda, M.P.C., Rizzo, R., Safanelli, J.L., Bonfatti, B.R., Silvero, N.E.Q. and Dematte, J.A.M., (2020). Soil color and mineralogy mapping using proximal and remote sensing in Midwest Brazil. Remote Sens. 12, 1197.
19 Ramos, P. V., Inda, A. V., Barrón, V., Siqueira, D. S., Júnior, J. M., & Teixeira, D. D. B. (2020). Color in subtropical brazilian soils as determined with a Munsell chart and by diffuse reflectance spectroscopy. *Catena*, *193*, 104609.
20 Rizzo, R., Wadoux, A. M. C., Demattê, J. A., Minasny, B., Barrón, V., Ben-Dor, E., Francos, N., Savin, I., poppiel, R., silvero,N. E. Q., Silva terra, F. D., Rosin, N. A., Rosas, J. T. F., Greschuk, L. T., Ballester, M. V. R., Das, B. S., Malone, B. P. & Salama, E. S. M. (2023). Remote sensing of the Earth's soil color in space and time. *Remote Sensing of Environment*, *299*, 113845.
21 Silvero, N. E. Q., Demattê, J. A. M., Amorim, M. T. A., dos Santos, N. V., Rizzo, R., Safanelli, J. L. &Bonfatti, B. R. (2021). Soil variability and quantification based on Sentinel-2 and Landsat-8 bare soil images: A comparison. Remote Sensing of Environment, 252, 112117.
22 ViscarraRossel, R.A., Bui, E.N., De Caritat, P. and McKenzie, N.J., (2010). Mapping iron oxides and the color of australian soil using visible-near-infrared reflectance spectra. J. Geophys. Res. Earth Surf. 115, 1–13.