JPMTR-2223
DOI 10.14622/JPMTR-2223
UDC 004.352-028.25-026.611|7.039

Original scientific paper | 176
Received: 2022-10-04
Accepted: 2023-04-24

An approach to predict print density using scanner and regression models

Shankhya Debnath[1] *and Arpitam Chatterjee*[2]

[1] Department of Printing Technology,
Regional Institute of Printing Technology, Kolkata, India
[2] Department of Printing Engineering, Jadavpur University, Kolkata, India

arpitam.chatterjee@jadavpuruniversity.in

Abstract

Optical density measurements are critical for process control and quality assurance during print production. Use of scanners and cameras for color correction and colorimetric measurements have been reported earlier in literature. This paper presents a procedure for using a scanner as a densitometer. Density of printed patches were calculated using the pixel intensity values obtained from scanning a test target. Optical density was measured simultaneously from a densitometer and also computed from the L^* measurements for the same patches. Four regression algorithms were used for modeling the behavior of the scanner using these two models. The models were tested and validated. The accuracy and performance of these models were compared. Results convey that the scanner can indeed be used for taking densitometric measurements obtained from the L^*, under the presented method.

Keywords: densitometer, process control, print quality, regression, machine learning

1. Introduction

Optical density is a measure of the share of light that gets reflected or transmitted from a surface. In the case of films, it is referred to as transmission density, while for opaque substrate, it is called reflection density. Process control plays a pivotal role in ensuring quality during print production. There exist multiple dimensions to the process control methodologies that are employed at press. Apart from colorimetric considerations, densitometric metrology and its control is required for achieving repeatable print results with other variables remaining the same. Many specifications and standards in print are based on achieving some pre-determined densitometric aims. Even when targeting the colorimetric aims, it is pertinent to achieve some standard print density. Density measurements provide a direct indication of the ink film deposition on the substrate. Offset printing is heavily reliant on the proportional quantities in which process inks are deposited. Optimal ink film deposition ensures good gray balance and printing without any color cast. Density can also give an insight into the quality of tonal reproduction during printing. In case of failure to attain densitometric targets or control thereof, problems that include hue shift during multi-color wet-on-wet printing are likely to occur. Densitometric measurements are also used for quantifying other parameters like dot gain, trap, contrast and hue error. Further, it is worthwhile to note that calibration of tonal values, which is a prerequisite to press characterization, also involves densitometric metrology.

The standard methodology for measurement of optical reflection density involves the use of a densitometer. Based on geometric construction and spectral behaviors of the instruments, there exists regional and international standards that define such measurement conditions. A densitometer typically employs a light source that is placed at an angle of 45° with respect to the sample where the light is focused using optical apparatus. The reflected light is recorded at an angle of 90°. This is done to capture the diffused light and ignore the gloss effects of the printed substrate. For traditional densitometers used in the printing industry, the reflected light would be made to pass through the color filters based on the ink whose density is being measured. For process colors, complimentary filters are used, while separate readings from each filter are required for non-process colors with the highest reading being the one corresponding to the density of such colors. However, present-day densitometers employ

spectral reflectance values at an optimum wavelength for the measurement of the densities of prints. This is especially important when measuring the densities of spot colors, as selecting any one of the fixed color filters would not yield optimal results. These instruments have built-in filter functions, each of which matches specific print condition (Eckhard, et al., 2014). The reflected light is then made to pass through a logarithmic amplifier and digitizer. The use of logarithmic scale is to ensure parity between densitometric measurements and logarithmic response nature of the human eyes. Further, to eliminate the differences in measurements owing to wet and dry ink films, polarization filters are provided both along the incident and reflected light path. The resulting reflected light is used to calculate the optical reflection density of the patch. Mathematically, optical density is given as the logarithm of the inverse of percentage reflectance.

Some of the earlier works (Derr, 1959) in constructing reflection densitometers have elaborated methods to develop the instrument to regulate the specular component of reflected light obtained from the printed surface. Kendall (1932) developed a reflection densitometer for the purpose of analysing the density of silver deposits on photographic papers. Watt (1956) developed a densitometer with similar geometry as referred to earlier, which was able to employ dense filters including interference filters to be used. McFarlane (1934) described a method for the construction of a versatile reflection densitometer with similar geometry that could be used for measurements on photographic papers as well as halftone printed substrate.

During print process control, faster densitometric measurements are required to implement any changes at press, if required. Densitometric measurements taken from patches printed on color bars are usually done from one patch to the other sequentially. This paper aims to implement a faster method in density measurements using an image scanner. Image scanners have been used in prior work in extracting densitometric measurements from plates (Lim and Mani, 1998; 1999), where the quality on solid prints referred to as mottle were evaluated from the lightness variation obtained from scanned image of a print. It is noteworthy that in the said work, the scanner was calibrated using calibration software and profiles were made for conversion of device dependent to independent data. In Wu (2001), investigation was conducted on the sources of error during measurements of print quality using a color scanner and proposed a method of linearizing followed by color scanner calibration to convert scanned colors from device space to independent space. In Rasmussen, Mishra, and Mongeon (2000), authors approached the problem of analysing print defects and quality using flatbed and drum scanners. They used multiple scanners for sample collection and converted the scanned color patches from RGB to CIELAB space using color transformation matrices. Metrics relating to lightness values of cyan, magenta, yellow and black patches printed on a test target were then utilized for print quality assessment. Scanners were also used for measuring print quality in Streckel, et al. (2003), where single color patches were scanned for analysing density followed by calibration of the scanner by means of converting the scanner responses to density using conversion tables.

Scanners have also been used in fields other than the graphic communication like in the medical technology field. Alva, et al. (2002), described a method wherein a flatbed image scanner was used to work like a densitometer for measuring the film response of radiochromic films. The films were irradiated in incremental manner leading to the formation of step wedges which were scanned and the film responses were evaluated and compared to optical density. An almost linear relationship was observed leading for the conclusion that film responses obtained from scanned films could be a substitute for optical density measurements. Xuong (1969) developed a system for measuring the density on x-ray diffraction films. The system used a drum type scanner for obtaining the measurements and the resultant light intensity were used for converting the film response to optical density. Hertel and Hultgren (2003) studied the granularity, which is a function of density differences, by measuring the densities from a scanned grayscale step-wedge on a flatbed scanner. Hertel, Töpfer and Böttcher (1994) used photodetector arrays to capture images of color films and as such calculated density from this setup. The results showed that the granularity density measured from the device and that obtained from a densitometer were comparable. In their study, Hertel and Brogan (2003) used a flat-bed scanner to analyze the image quality of prints. They processed the RGB signals to visual density measurements using visual weighing coefficients. A method for using a stepless wedge with varying density and a flatbed scanner to measure granularity versus density changes was developed by Hertel and Hultgren (2002). Brydges, et al. (1998), described a method of using a CCD color camera for measuring raw RGB values of printed patches to obtain densitometric measurements from the readings. Seymour (1995) has used a CCD camera for on-line detection of quality in printed output. The work proposes to use the measured grayscale values for R, G, B components of a patch obtained from the camera and their corresponding densities obtained from a densitometer to develop a look-up table and use the same for further calculation of densities of unknown patches. Other authors (Simomaa, 1987; Nemeth and Wang, 1993; Malmqvist, et al., 1993; Busk, et al., 1993) have also used CCD cameras for measuring optical densities of print.

In this paper, a method is presented of using a flatbed scanner for measuring the density of printed substrate using machine learning. There have been wide applications of machine learning techniques in identifying and analysing print quality. Verikas and Bacauskiene (2008) have demonstrated the use of a CCD-based color camera to obtain the spectral reflectance of color patches and converted these measurements to ink film density using a local kernel ridge regression method. Lundström, et al. (2013), have used a CCD-based camera for scanning printed patches on a web offset machine and implemented a random forest based algorithm to estimate ink density among other quality parameters. Al-Mutawa and Moon (1993) have used a connectionist expert system that learns the relationship between changes made by the press operator in ink key settings based on changes in the ink film density and automates the process. In Yang, Chou and Yang (2013), the authors used support vector regression (SVR) to calibrate non-linear systems like camera or computer for correct color reproduction. Kuo, Ng and Wang (2002) used SVM to compute differential gloss on printed substrate from a mapping function obtained from the density/gloss measurements. Verikas, Bacauskiene and Nilsson (2006) used SVR among other soft computing techniques to evaluate values of various print quality attributes to ensure print quality during production. Funt and Xiong presented a method of using SVR in determining the chromaticity of the light that has been used in illuminating a scene from the histogram of the image (Funt and Xiong, 2004; Xiong and Funt, 2006). Evans and Fisher (1994) used decision-tree modelling that resulted in reduction of banding during gravure print production. Das, et al. (2022), have presented a method of using various machine learning including decision tree to solve the delays in print production of gravure printing that involves cylinder banding. Rabiha, et al. (2018), used decision tree based data-mining methods to develop a set of rules and segment of customers that helped a printing company to push marketing strategies.

There have been multiple implementations of regression methods to solve problems pertaining to print and print quality. Various methodologies towards scanner and other optical devices characterization have reportedly used these techniques. Izadan and Nobbs (2006) have used regression techniques for modelling the scanner behaviour in converting the RGB values to XYZ space. Lundström and Verikas (2013) used regression techniques among other methods to quantify and assess quality of print based on various parameters. Multiple linear regression methods were used for characterizing the colorimetric behaviour of desktop scanners while developing a color management module for desktop-based printing system in Iino and Berns (1998). In their work, Gebejes, et al. (2013), showed methods for recovering reflectance data from multi-angular camera RGB data using regression methods. Kucuk, et al. (2022), have studied and compared the results of camera color correction for converting camera RGB data to XYZ data based on regression method and that obtained from neural network based method. Bangyong, Han and Shisheng (2014) applied polynomial regression methods to model the relationship between the measured and printed colors of patches in a test target to ultimately obtain the gray values of the patches. Hirn, et al. (2009), have reported the relationship between print density differences and the local properties of the paper and modelled this using linear regression method. A method of calibrating scanner for converting recorded device-dependent color data to independent one using polynomial regression has been proposed by Hardeberg, et al. (1996). Another method of using polynomial regression for characterizing digital color camera by converting native space color data to independent data was proposed by Hong, Luo and Rhodes (2001) and Han (1998). Shaw, et al. (2003), and Yang, et al. (2010), have used principal component regression in characterizing printers and modelling their output behaviour. Jetsu, et al. (2006), and Lo, et al. (2006) presented methods for characterizing printers and converting device-dependent color to their corresponding reflectance spectra using polynomial regression, among others.

The contribution of this paper includes a machine learning based framework for prediction of densitometric measurement from the scanned color patches and a comparative analysis between performances of some of the popular conventional prediction techniques on the same. The success of the presented method can result in development of much less expensive and easy to operate system for density measurement since the commercially used densitometers are considerably expensive and often not available in small presses, while scanner is available in most of the presses.

2. Methods

Prior to the experimentation, a test target was developed. Since the work involved characterizing the behaviour of a scanner, the test target was made in line with IT 8.7/2. The patches were designed with CIE $L^*a^*b^*$ values from a reference file obtained online. Since this target is a standard that is used worldwide, it was seen fit to use it for the work. The target contains 264 color patches and 24 gray patches. Once the target was developed, it was printed using a CMYK inkjet desktop printer on an optical brightening agent (OBA) free photopaper. The target was then scanned using an Epson Perfection V19 Photo scanning device.

Since any changes that might be done to the image during scanning or processing might lead to alteration of pixel intensity values of the image, and it might lose its fidelity to the original, the scan was done in RAW mode, without any further processing like color correction, sharpening and exposure control. The image was scanned with linear Gamma at 8-bit RGB color and 1200 dpi resolution. The scanned image was stored in uncompressed TIFF format.

The image was analyzed using ImageJ (Pérez and Pascau, 2013). The mean pixel intensity values for individual patches were extracted using the software from the scanned image of the target. Since optical density is a measure of the logarithmic ratio of incident and reflected light (Merton, 1924), the incident light (I_0) has been considered here as the highest possible pixel intensity value for an 8-bit RGB image, i.e., 255, while the reflected light (I) is mean pixel intensity obtained for a patch, as given in Equation [1].

$$\mathrm{OD} = \log_{10}\left(\frac{I_0}{I}\right) \qquad [1]$$

The mean pixel intensity was extracted from the individual patches, and the optical density (OD) was derived for the patches using Equation [1].

For characterizing the behaviour of the scanner, two models were developed. The first was comparing the densities obtained from mean pixel intensities (OD) of the patches to their corresponding actual optical densities (MD) obtained from a densitometer. Hence, for this model, OD was the independent variable and MD was the dependent variable. This will be further referred to as the MD model. The second model involved extracting the L^* values of the said patches using a spectrophotometer. The L^* refers to the lightness of luminance component (León, et al., 2006) of a color and is chroma-independent. The CIE $L^*a^*b^*$ color space is perceptually uniform and as such it is a metric that can define the darkness or lightness of a patch. A Techkon SpectroDens spectrophotometer was used for measuring the L^* values for the patches in the test target. The M1 measurement conditions were used with D50/10° illuminant/observer. Equations [2] and [3] describe the method to compute L^* values.

$$L^* = 116\left(\frac{Y}{Y_n}\right)^{\frac{1}{3}} - 16, \text{ if } \left(\frac{Y}{Y_n}\right) \geq 0.008856 \qquad [2]$$

$$L^* = 903.3\left(\frac{Y}{Y_n}\right), \text{ if } \left(\frac{Y}{Y_n}\right) < 0.008856 \qquad [3]$$

where L^* is the CIE lightness value (Ebner, 2007), and Y_n is the Y tristimulus value for the reference white used for measurement. The Y tristimulus value represents the percentage luminance factor. The ratio Y/Y_n when raised to the one third cube represents a perceptual attribute of lightness (Sharma, 2018). The value of Y is 100 for perfect white patches and 0 for patches that absorb no light (Hunt and Pointer, 2011). Thus, this metric can be used for converting the L^* values to optical density, as the ratio (Y/Y_n) correlates to the ratio (I/I_0) as given in Equation [1]. This results in the following equations:

$$\mathrm{LD} = \log_{10}\left(\frac{116}{L^* + 16}\right)^3, \text{ if } \frac{Y}{Y_n} \geq 0.008856 \qquad [4]$$

$$\mathrm{LD} = \log_{10}\left(\frac{903.3}{L^*}\right), \text{ if } \frac{Y}{Y_n} < 0.008856 \qquad [5]$$

Equations [4] or [5], depending on satisfying the given condition, were used for converting the L^* values to optical density (LD) for the measured patches. These are the actual density values for the patches (dependent variable) that were used along with OD (independent variable) for modelling the behaviour of the scanner. This will be further referred to as the LD model. An example of scanned patches and corresponding mean intensity and L^* values is shown in Figure 1.

Along with the density-based model (MD), the L^* based model (LD) was also considered owing to two reasons: L^* is an indicator of darkness or lightness of a patch and having two distinct models for prediction would allow scope of cross-validation of results. For both the models, i.e., MD and LD, four regression techniques were applied to find the best prediction model; namely LR: linear regression, PR-2: second-order polynomial regression, SVR: support vector regression, and DT: decision trees. A brief of these algorithms has been provided below for ready reference. In this work, for

	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20	21	22
Mean pixel intensity	35.00	38.81	42.67	45.19	66.33	72.56	76.76	77.29	127.88	133.36	139.45	137.02	203.55	206.46	203.73	200.42	198.97	201.52	202.76	25.22	41.62	58.91
L^* value	17.17	18.23	19.91	20.97	32.15	33.57	34.58	33.95	58.33	59.76	60.63	58.63	79.49	79.52	79.46	79.57	79.71	80.18	80.02	10.89	18.58	24.84

Figure 1: Example of scanned printed patches and corresponding mean intensity and L values*

Table 1: Pseudocodes of used regression algorithms

Linear and polynomial regression	Support vector regression	Decision trees
General form of representation: $y = a_0 + a_1x + a_2x^2 \ldots + a_nx^n$ For linear regression $n = 1$ Run the algorithm to find the coefficients (a_0, a_1, a_2, …, a_n) following the least square method where the sum of squared residuals (S) is calculated as below and minimized. $S = \sum_{i=1}^{n} r_i^2$ where $r_i = y_i - f(x_i, a_i)$	Unlike linear and polynomial regression the goal is to minimize the L2 norm of the coefficients represented as: $\min \frac{1}{2}\|a\|^2$ where the error term is included in terms of a constrained presented as $\|y_i - a_i x_i\| \leq \varepsilon$ where ε is the maximum acceptable error.	This algorithm learns the decision rules from the training data. It starts with a root node and then keep on splitting into child node in order to reduce entropy (E) and increase information gain (IG) that are calculated as follows: $E = \sum_{i=1}^{N} - p_i \log_2(p_i)$ $\mathrm{IG} = E_l - \sum_{j=1}^{K} E_{j,l+1}$ where l is the iteration and j is the index of subject in total number of subset K in that split.

all the models, 60 % of the total number of patches were used for training, 20 % for model validation and remaining 20 % for testing.

The stated regression algorithms were subjected to 10-fold cross-validation to arrive into the model parameters to be used for testing. Among these algorithms DT needs to have two important user specified parameters, i.e., maximum depth of tree, which regulates the maximum number of splits a tree can make, and minimum number of sample leaf, which conveys the minimum number of samples or observations needed to make a leaf under a parent node. From the literature review (Doğru, Buyrukoğlu and Arı, 2023; Hou, et al., 2023; Yazu, et al., 2022) and trial- and-error runs with different combinations of values, it has been found that 40 for depth of tree and 10 for minimum sample leaf may be optimal for this work. Nevertheless, scope of optimizing the DT parameters is still open and has not been considered under the focus of this work. The LD and MD based models were then assessed based on their accuracy and error metrics.

2.1 Regression algorithms

Regression algorithms are commonly used for prediction of continuous real values. Regression analysis corresponds to the mathematical process of identifying the relationship between a dependent variable and one or more independent variable. Among different algorithms, linear and polynomial regression (Weisberg, 2005; Seber, 2012; Ostertagová, 2012), support vector regression, (Vapnik, Golowich and Smola, 1997; Vapnik, 1999; Drucker, et al., 1996), and decision trees (Kingsford and Salzberg, 2008; Kotsiantis, 2013; Breiman, et al., 1984) are the popular algorithms that have been explored in this work. The pseudocodes of these algorithms are shown in Table 1.

3. Results

The results of 10-fold cross validation runs have been provided in Table 2, where the best folds have been highlighted in bold.

Table 2 shows that LR, SVR and DT perform equivalently while PR-2 shows better performance in comparison to other algorithms under consideration. The betterment is reflected in the average and standard deviation values as well. In case of PR-2, the average value is higher than others and standard deviation value is lower. Table 2 also shows that LD model provides visibly better prediction performance than the MD model for all the regression models under consideration.

Figures 2 and 3 show the prediction plots of the MD model and LD model, respectively, for the regression models under consideration. In these plots the diagonal line presents ideal line where predicted values exactly match the actual values. Hence, the deviation of the scattered points from the diagonal line shows the goodness of the prediction.

The accuracy of prediction can further be visualized using residual plots shown in Figure 4 and Figure 5 for the MD based and LD based models, respectively. In these plots the differences in percentage between actual and predicted values have been plotted. The differences between predicted and actual values were

Table 2: Results of 10-fold cross-validation for different regression algorithms for both MD and LD models

	LR		**PR-2**		**SVR**		**DT**	
	MD	LD	MD	LD	MD	LD	MD	LD
Fold 1	**0.8781**	0.9555	0.8721	0.9855	0.8683	0.9855	0.7610	**0.9855**
Fold 2	0.6435	0.9805	0.6835	0.9555	0.6541	0.9555	0.7090	0.9555
Fold 3	0.7600	0.9627	0.7900	0.9805	0.7550	0.9805	0.7742	0.9505
Fold 4	0.6302	0.9792	0.8172	0.9627	0.6494	0.9627	0.5840	0.9627
Fold 5	0.7562	0.9642	0.8232	0.9792	0.7506	0.9792	0.6978	0.9592
Fold 6	0.8476	0.9743	0.8500	**0.9912**	**0.8832**	0.9642	**0.8149**	0.9642
Fold 7	0.7229	**0.9866**	0.8120	0.9743	0.6978	0.9743	0.7474	0.9743
Fold 8	0.7238	0.9715	0.7842	0.9866	0.7038	**0.9866**	0.6779	0.9766
Fold 9	0.8137	0.9521	**0.8847**	0.9715	0.7894	0.9715	0.7882	0.9515
Fold 10	0.6358	0.9555	0.7829	0.9521	0.6309	0.9521	0.7386	0.9521
Average	0.7411	0.9682	0.8099	0.9739	0.7382	0.9712	0.7293	0.9632
Standard deviation	0.0878	0.0113	0.0568	0.0126	0.0883	0.0115	0.0661	0.0113

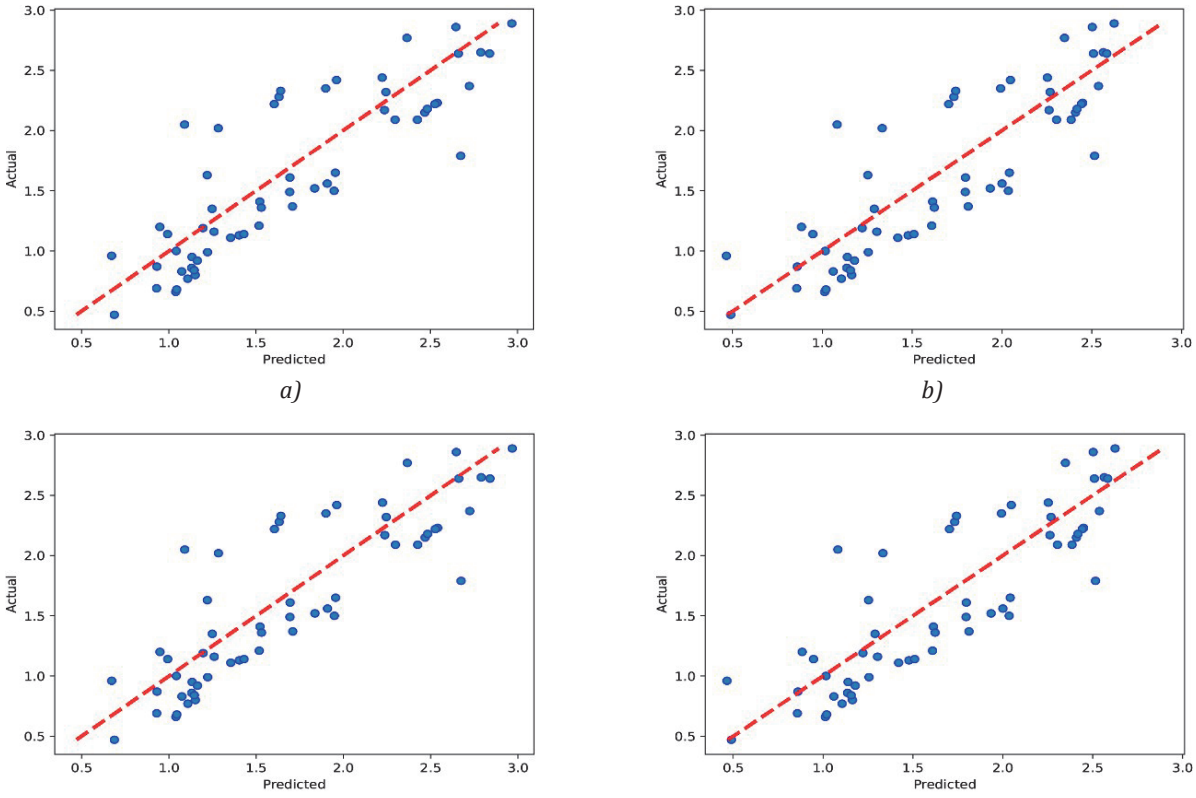

Figure 2: Prediction results with the MD model for different regression algorithms; (a) LR, (b) PR-2, (c) SVR and (d) DT

interpreted in percentage values using Equation [6]. The straight line at '0' is the reference line that would result if the predicted values were exactly same as the actual values.

$$\text{Residual value (\%)} = \frac{|x_i - y_i|}{x_i} \times 100 \qquad [6]$$

where x_i and y_i are individual actual and predicted values, respectively.

Apart from the visual presentations, the results of predictions were further evaluated against some of the popular metrics (Naser, 2020; Ostertagová, 2012), namely, mean squared error (MSE), mean absolute error (MAE) and *R*-squared (R^2) values. The first two metrics represent goodness of the prediction model by small values, while higher the R^2 values the better prediction performance. The results of evaluation for the MD and LD models are presented in Table 3.

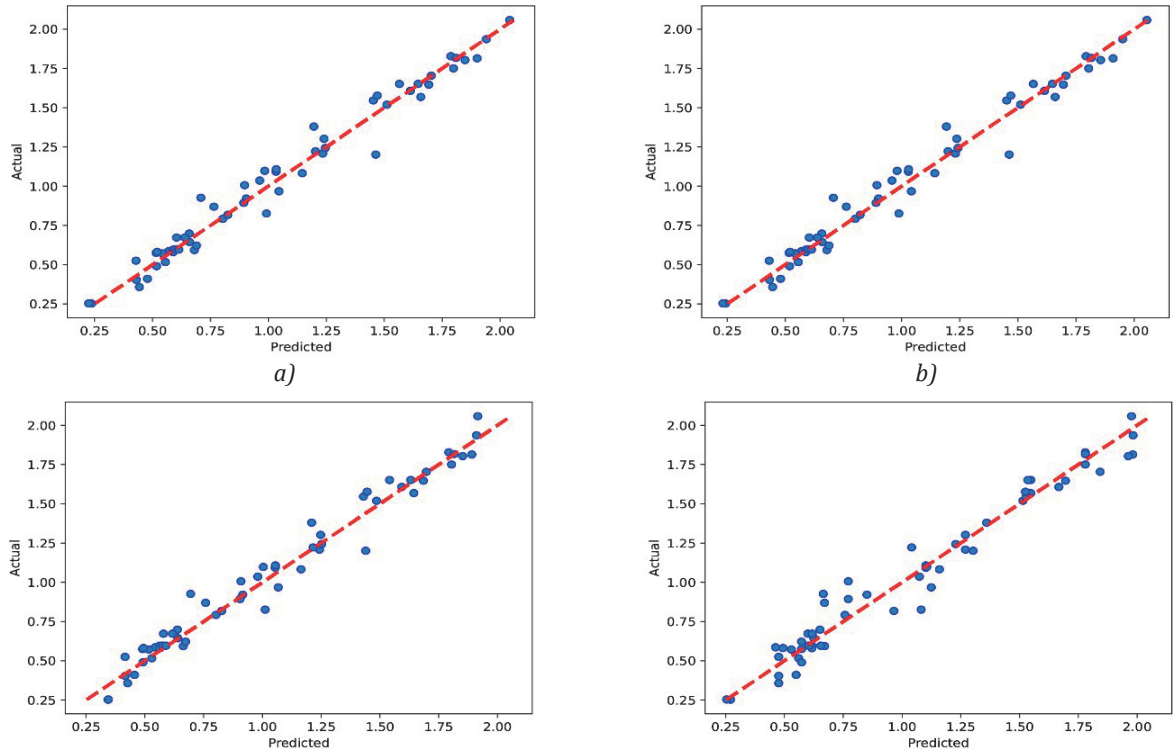

Figure 3: Prediction results with the LD model for different regression algorithms; (a) LR, (b) PR-2, (c) SVR and (d) DT

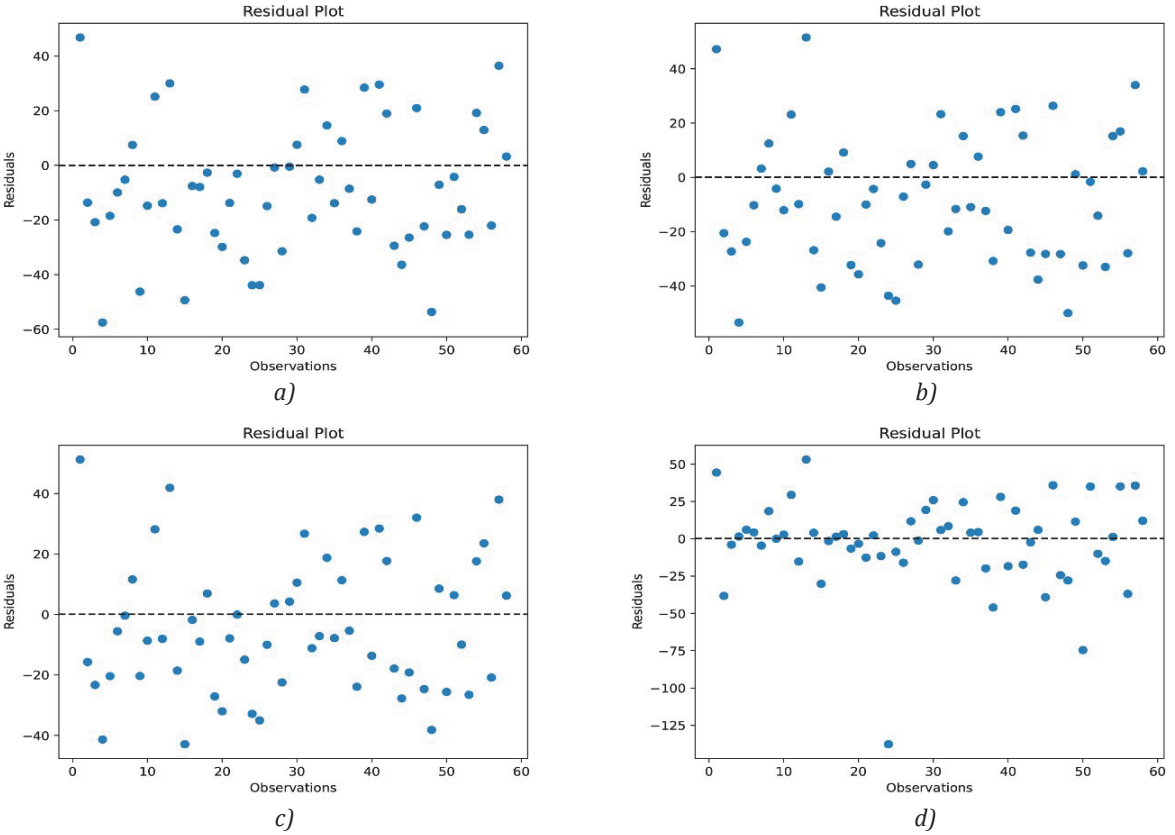

Figure 4: The residual plots with the MD model for different regression algorithms; (a) LR, (b) PR-2, (c) SVR and (d) DT

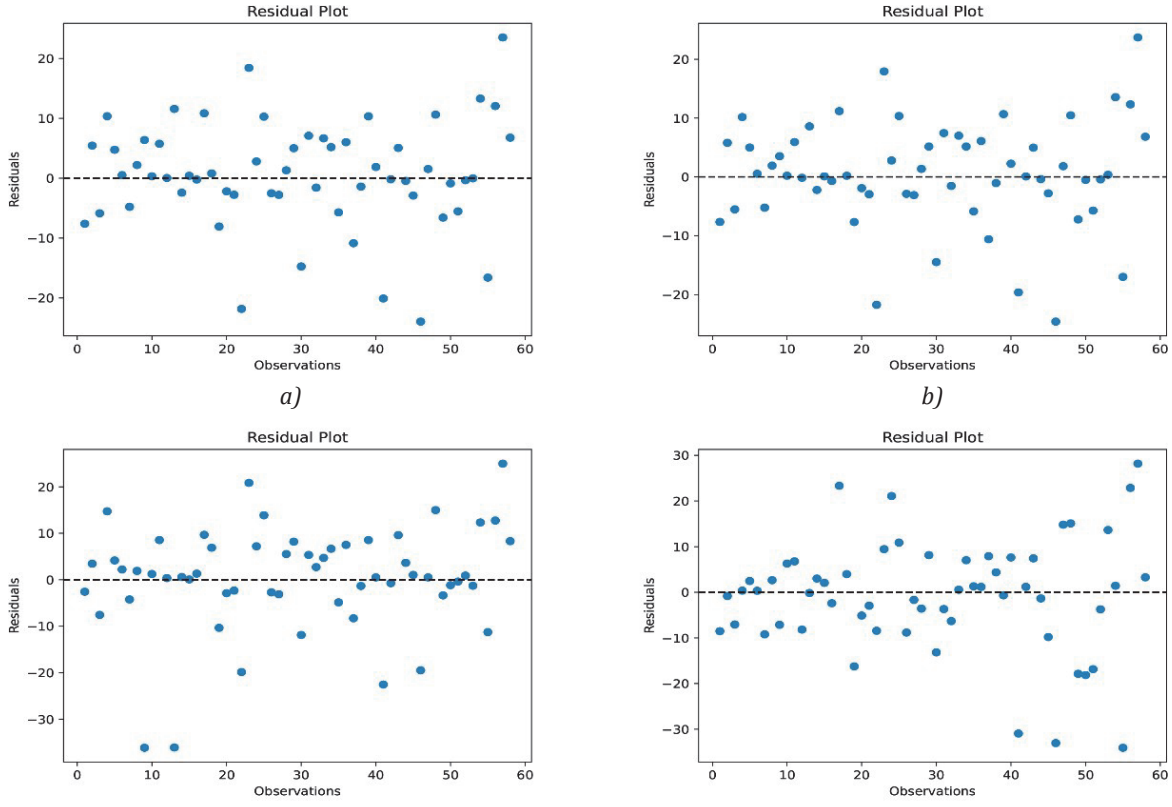

Figure 5: The residual plots with the LD model for different regression algorithms; (a) LR, (b) PR-2, (c) SVR and (d) DT

Table 3: Performance evaluation of the models with various regression algorithms

Regression algorithm	**MSE**		**MAE**		**R^2**	
	MD	LD	MD	LD	MD	LD
LR	0.1261	0.0060	0.2967	0.0559	0.7204	0.9753
PR-2	0.1235	0.0013	0.2998	0.0256	0.7262	0.9932
SVR	0.1188	0.0069	0.2779	0.0627	0.7366	0.9718
DT	0.1423	0.0098	0.2839	0.0756	0.6845	0.9597

4. Discussions

From the MD model prediction results that can be seen in Figure 2, it is clearly evident that the predicted points are widely spread out away from the diagonal line, indicating poorer goodness of fit. From the prediction results of all four regression algorithms, it is further evident that the dependent (MD) and independent (OD) variables most likely do not share a linear relationship.

The prediction results shown in Figure 3 for the LD model show that in all the cases the predicted points are in close proximity to the diagonal line, which reflects the goodness of fit. The figure also conveys the low degree of relational non-linearity between OD and LD. Due to such relation, simple regression models like LR and PR-2 have also shown good prediction results.

Figure 4 describes the residual plots for the MD model. The residues are again more widely spread out for all the regression algorithms as compared to the LD model. The overall spread for the residues has increased as can be seen from the increased limits for the axes that vary within ±45 for LR, PR-2 and SVR algorithms and from +55 to –150 for the DT algorithm, indicating poor predictors.

Residual plots in Figure 5 for the LD model show that residues are denser near the origin, scarce away from it and have no apparent patterns when moved along the *x*-axis, which are characteristics of good plots. The residual plots validate the model performance and increase the confidence in its accuracy. The plots for PR-2, LR and SVR suggest that most of the data resides within ±10 % from the origin. There is no obvious patterns observable in the residuals. The plot for DT model contained more residuals away from the origin in comparison to other models. And it is quite noticeable that the residuals for PR-2 are concentrated along the origin line more than for any other model. This is in line with the observations from the earlier findings. From the results discussed, it may be understood that the PR-2 model performed better than other models for this experiment.

Table 3 shows that in terms of the accuracy of the four regression algorithms, the PR-2 model with the R^2 error at 0.9932 performs the best among all. The results for MSE and MAE also are the lowest for the PR-2 model. This can be verified with the fact that the independent variable is one-dimensional and is highly correlated with the dependent variable. The R^2 error for PR-2 and LR are similar, yet MSE and MAE values for PR-2 are significantly higher than for LR. It is interesting to note that performance of models LR and SVR is almost comparable. It may be due to the fact that SVR performs well for multidimensional input variables. It seems that DT has performed relatively poorer in modeling the scanner behavior. The reason might be attributed to the fact that DT results in over-fitting frequently, hence there are comparatively more cases where DT could not perform well in predicting unknown observations. However, if their variables had higher degree of non-linearity, DT could have performed well compared to LR and PR-2. Although the R^2 values are indicative of the accuracy of the model, other metrics are required to validate the models. Residual plot might be useful in visualizing and validating the performance of the regression algorithms and MAE and MSE values may act as the validator for R^2 results.

Table 3 also presents the error metrics for MD model. As evident, the error metrics for all the regression algorithms have poorer values when compared to those obtained from the LD model. Of the four regression algorithms used for the MD model, PR-2 performed better than the rest. When compared to their LD counterparts, the MD regression algorithms fell short by more than 24 % in R^2 error, by more than 73 % in MAE, and by more than 93 % in MSE. This finding is in line with the prediction and residual data discussed earlier.

When compared to the MD model, the LD model certainly not only outperformed it but also predicted results with substantial accuracy. Overall, the results are encouraging and lead to the inference that using the PR-2 regression algorithm of the LD model, the scanner behavior can be characterized and the scanner as such can be used for the purpose of densitometric measurements.

5. Conclusion

This work provides an insight into a new regression-based investigation that may be used for characterizing a scanner for using it as a densitometer. Two models were tested, one based on actual density measurements and the other based on density measurements obtained from L^* values. Multiple regression algorithms were implemented for both these models in data analysis and the results suggested that PR-2 algorithm of the LD model outperformed other machine learning algorithms. One reason behind this might be the fact that the independent input variables were one-dimensional and machine learning algorithms with higher computational complexity deal well with complex multidimensional data. The scope of this work included estimation of density of patches printed using inkjet technology on a single type of paper. While working with photopapers, the measurement condition is important. It might be prudent to use the polarizing filter along with M2 measurement condition for substrates that are glossy and fluorescent. Further work may include measuring density on a wide variety of substrates printed with multiple technologies and collecting other data along with pixel intensities. In such cases, machine learning algorithms might prove efficient in handling larger multi-variate data. Instead of measuring the density of individual patches, a single scan can provide the density of hundreds of patches, resulting in increased efficiency of process control during print production. The reported work thus can be a promising step towards mapping scanned print patches into possible densitometric measurements.

Acknowledgement

The authors would like to express their sincere gratitude to Mr. Manu Choudhury, Director, CDC Printers, Kolkata, India for allowing them to use his spectrophotometer for taking the measurements of the test targets.

References

Al-Mutawa, S. and Moon, Y.B., 1993. Process drift control in lithographic printing – issues and a connectionist expert system approach. *Computers in industry*, 21(3), pp. 295–306. https://doi.org/10.1016/0166-3615(93)90026-W.

Alva, H., Mercado-Uribe, H., Rodríguez-Villafuerte, M. and Brandan, M.E., 2002. The use of a reflective scanner to study radiochromic film response. *Physics in Medicine & Biology*, 47(16): 2925. https://doi.org/10.1088/0031-9155/47/16/308.

Bangyong, S., Han., L. and Shisheng, Z., 2014. Calculating cyan-magenta-yellow-black (CMYK) printer gray component data based on polynomial modeling. *Scientific Research and Essays*, 9(9), pp. 352–356. https://doi.org/10.5897/SRE2014.5915.

Breiman, L., Friedman, J.H., Olshen, R.A. and Stone, C.J., 1984. *Classification and regression trees.* Belmont, CA, USA: Wadsworth International Group.

Brydges, D., Deppner, F., Kunzli, H., Heuberger, K. and Hersch, R.D., 1998. Application of a 3-CCD color camera for colorimetric and densitometric measurements. In: *Proceedings SPIE 3300, Color imaging: Device-Independent Color, Color Hardcopy, and Graphic Arts III*. San Jose, CA, USA, 24–30 January 1998. SPIE. https://doi.org/10.1117/12.298292.

Busk, H., Malmqvist, L., Malmqvist, K. and Bergman, L., 1993. Image analysis for the development of multicolour print quality in newspaper printing. In: W.H. Banks, ed. *Advances in Printing Science and Technology: Proceedings of 22nd Research Conference of the International Association of Research Institutes for the Graphic Arts Industry*. Munich, Germany, 5–8 September 1993. London: Pentech Press.

Das, A., Rakshit, P., Roy, S., Dutta, B.R., Ghosh, A. and Mitra, D., 2022. Rotogravure printing band analysis with the help of machine learning. In: J.K. Mandal, M. Hinchey, S., Sen and P. Biwas, eds. *Applications of Networks, Sensors and Autonomous Systems Analytics*. Kalyani, India, 11–12 December 2020. Singapore: Springer. https://doi.org/10.1007/978-981-16-7305-4_20.

Derr, A.J., 1959. Optical unit for reflection densitometry. *Journal of the Optical Society of America*, 49(2), pp. 176–178. https://doi.org/10.1364/JOSA.49.000176.

Doğru, A., Buyrukoğlu, S. and Arı, M., 2023. A hybrid super ensemble learning model for the early-stage prediction of diabetes risk. *Medical & Biological Engineering & Computing*, 61(3), pp. 785–797. https://doi.org/10.1007/s11517-022-02749-z.

Drucker, H., Burges, C.J.C., Kaufman, L., Smola, A.J. and Vapnik, V., 1997. Support vector regression machines. In: M.C. Mozer, M. Jordan and T. Petsche, eds. *Advances in Neural Information Processing Systems 9 (NIPS 1996).* Denver, CO, USA, 2–5 December 1996. The MIT Press.

Ebner, M., 2007. *Color constancy*. Chichester, UK: John Wiley & Sons.

Eckhard, T., Klammer, M., Valero, E.M., and Hernández-Andrés, J., 2014. Improved spectral density measurement from estimated reflectance data with kernel ridge regression. In: A. Elmoataz, O. Lezoray, F. Nouboud and D. Mammass, eds. *Image and Signal Processing: 6th International Conference, ICISP 2014*. Cherbourg, France, 30 June – 2 July 2014. Cham: Springer International Publishing. https://doi.org/10.1007/978-3-319-07998-1_10.

Evans, B. and Fisher, D., 1994. Overcoming process delays with decision tree induction. *IEEE Expert*, 9(1), pp. 60–66. https://doi.org/10.1109/64.295130.

Funt, B. and Xiong, W., 2004. Estimating illumination chromaticity via support vector regression. In: *Proceedings of the Twelfth Color and Imaging Conference*. Scottsdale, AZ, USA, 7–11 November 2004. Society for Imaging Science and Technology, pp. 47–52.

Gebejes, A., Martinez Domingo, M.Á, Heikkinen, V. and Tomic, I., 2013. Reflectance recovery for coated printed color samples via multiangular RGB camera measurements. In: J.Y. Hardeberg and M. Pedersen, eds. *Proceedings of 2013 Colour and Visual Computing Symposium*. Gjøvik, Norway, 5–6 September 2013. IEEE. https://doi.org/10.1109/CVCS.2013.6626287.

Hardeberg, J.Y., Schmitt, F., Tastl, I., Brettel, H. and Crettez, J.-P., 1996. Color management for color facsimile. In: *Proceedings of 4th IS&T/SID Color Imaging Conference*. Scottsdale, AZ, USA, 19–22 November 1996. Society for Imaging Science and Technology, pp. 108–113. https://doi.org/10.2352/CIC.1996.4.1.art00030.

Hertel, D.W. and Brogan, J.G., 2003. Polaroid scanner-based image quality measuring system. In: *Proceedings of PICS Conference: The Digital Photography Conference*. Rochester, NY, USA, 13 May 2003. Society for Imaging Science and Technology, pp. 140–146.

Hertel, D.W. and Hultgren, B.O., 2002. Scanner-based granularity measurement on a continuous density wedge. In: *Proceedings of 18th International Conference on Digital Printing Technologies*. San Diego, CA, USA, 29 September – 4 October 2002. Society for Imaging Science and Technology, pp. 189–194.

Hertel, D.W. and Hultgren, B.O., 2003. One-step measurement of granularity versus density, graininess, and micro-uniformity. In: *Proceedings of PICS Conference: The Digital Photography Conference*. Rochester, NY, USA, 13 May 2003. Society for Imaging Science and Technology, pp. 552–557.

Hertel, D., Töpfer, K. and Böttcher, H., 1994. Image quality investigations by means of photodetector arrays. *Journal of Imaging Science and Technology*, 38(1), pp. 44–48.

Hirn, U., Lechthaler, M., Wind, E. and Bauer, W., 2009. Linear regression modelling of local print density in gravure printed SC paper. In: *Papermaking Research Symposium – CD ROM Proceedings*. Kuopio, Finland, 1–4 June 2009.

Hong, G., Luo, M.R. and Rhodes, P.A., 2001. A study of digital camera colorimetric characterization based on polynomial modeling. *Color Research and Application*, 26(1), pp. 76–84. https://doi.org/10.1002/1520-6378(200102)26:1<76::AID-COL8>3.0.CO;2-3.

Hou, S., Liu, Y., Zhuang, W., Zhang, K., Zhang, R. and Yang, Q., 2023. Prediction of shield jamming risk for double-shield TBM tunnels based on numerical samples and random forest classifier. *Acta Geotechnica*, 18(1), pp. 495–517. https://doi.org/10.1007/s11440-022-01567-9.

Hunt, R.W.G. and Pointer, M.R., 2011. *Measuring colour*. 4th ed. Chichester: John Wiley & Sons.

Iino, K. and Berns, R.S., 1998. Building color-management modules using linear optimization I. Desktop color system. *Journal of Imaging Science and Technology*, 42(1), pp. 79–94. https://doi.org/10.2352/J.ImagingSci.Technol.1998.42.1.art00010.

Izadan, H. and Nobbs, J.H., 2006. Input device characterisation: a comparison between iteration and regression methods using either XYZ or L*a*b*. In: *Proceedings of CGIV 2006, 3rd European Conference on Color in Graphics, Imaging, and Vision*. Leeds, UK, 19–22 June 2006. Society for Imaging Science and Technology, pp. 158–162.

Jetsu, T., Heikkinen, V., Parkkinen, J., Hauta-Kasari, M., Martinkauppi, B., Lee, S., Ok, H. and Kim, C.Y., 2006. Color calibration of digital camera using polynomial transformation. In: *Proceedings of IS&T CGIV 3rd European Conference on Colour in Graphics, Imaging, and Vision*. Leeds, UK, 19–22 June 2006. Society for Imaging Science and Technology, pp. 163–166. https://doi.org/10.2352/CGIV.2006.3.1.art00032.

Kendall, C.W., 1932. A reflection densitometer for photographic papers. *Review of Scientific Instruments*, 3(11), 668–674. https://doi.org/10.1063/1.1748883.

Kingsford, C. and Salzberg, S.L., 2008. What are decision trees? *Nature Biotechnology*, 26(9), pp. 1011–1013. https://doi.org/10.1038/nbt0908-1011.

Kotsiantis, S.B., 2013. Decision trees: a recent overview. *Artificial Intelligence Review*, 39(4), pp. 261–283. https://doi.org/10.1007/s10462-011-9272-4.c

Kucuk, A., Finlayson, G.D., Mantiuk, R. and Ashraf, M., 2022. Comparison of regression methods and neural networks for colour corrections. In: *London Imaging Meeting 2022*. London, UK, 6–8 July 2022. Society for Imaging Science and Technology, pp. 74–79.

Kuo, C., Ng, Y. and Wang, C.J., 2002. Gloss patch selection based on support vector regression. In: *PICS 2002 Image processing, Image Quality, and Image Capture Systems Conference Proceedings*. Portland, OR, USA, 7–10 April 2002. Society for Image Science and Technology, pp. 121–125.

León, K., Mery, D., Pedreschi, F. and León, J., 2006. Color measurement in *L*a*b** units from RGB digital images. *Food Research International*, 39(10), pp. 1084–1091. https://doi.org/10.1016/j.foodres.2006.03.006.

Lim, W. and Mani, S., 1998. Application of digital imaging to measure print quality. In: *Proceedings of 14th International Conference on Printing Technologies (NIP 14)*. Toronto, Canada, 18–23 October 1998. Society for Imaging Science and Technology, pp. 611–614.

Lim, W. and Mani, S., 1999. Application of digital image analyses to measure print quality. *Journal of Coatings Technology*, 71(894), pp. 73–78. https://doi.org/10.1007/BF02698374.

Lo, M.-C., Chen, C.-L., Perng, R.-K. and Hsieh, Z.-X. 2006. The characterisation of colour printing devices via physical, numerical and LUT models. In: *Proceedings of CGIV 2006, IS&T's Third European Conference on Color in Graphics, Imaging, and Vision*. Leeds, UK, 20–22 June 2006. IS&T, pp. 95–99.

Lundström, J. and Verikas, A., 2013. Assessing print quality by machine in offset colour printing. *Knowledge-Based Systems*, 37, pp. 70–79. https://doi.org/10.1016/j.knosys.2012.07.022.

Lundström, J., Verikas, A., Tullander, E. and Larsson, B., 2013. Assessing, exploring, and monitoring quality of offset colour prints. *Measurement*, 46(4), pp. 1427–1441. https://doi.org/10.1016/j.measurement.2012.11.037.

Malmqvist, K., Bergman, L., Busk, H. and Malmqvist, L., 1993. The 3-colour CCD camera as a densitometer for measuring density of cyan, magenta and yellow in printed solid areas and in screen areas. In: W.H. Banks, ed. *Advances in Printing Science and Technology: Proceedings of 22nd Research Conference of the International Association of Research Institutes for the Graphic Arts Industry*. Munich, Germany, 5–8 September 1993. London: Pentech Press.

McFarlane, J.W., 1934. A reflection densitometer. *Journal of Optical Society of America*, 24(1), pp. 19–24. https://doi.org/10.1364/JOSA.24.000019.

Merton, T.R., 1924. On ultra-violet spectro-photometry. *Proceedings of the Royal Society A*, 106(738), pp. 378–384. https://doi.org/10.1098/rspa.1924.0076.

Nemeth, R. and Wang, B., 1993. Applying video technology to color measurement for the graphic arts. In: *TAGA 45th Annual Technical Conference Proceedings*. Minneapolis, MN, USA, 25–28 April 1993. Technical Association of the Graphic Arts, pp. 445–461.

Ostertagová, E., 2012. Modelling using polynomial regression. *Procedia Engineering*, 48, pp. 500–506. https://doi.org/10.1016/j.proeng.2012.09.545.

Pérez, J.M.M. and Pascau, J., 2013. *Image processing with ImageJ*. Birmingham: Packt Publishing.

Rabiha, S.G., Murmanto, I.R., Sasmoko, S., Yossy, E. and Kusumastuti, D.L., 2018. Consumer segmentation using case based reasoning approach to printing company. In: *2018 International Seminar on Research of Information Technology and Intelligent Systems (ISRITI)*. Yogyakarta, Indonesia, 21–22 November 2018. IEEE, pp. 327–331. https://doi.org/10.1109/ISRITI.2018.8864372.

Rasmussen, R., Mishra, B. and Mongeon, M.C., 2000. Using drum and flatbed scanners for color image quality measurements. In: *IS&T's PICS 2000: Image Processing, Image Quality, Image Capture, Systems Conference Proceedings*. Portland, OR, USA, 26–29 March 2000. Society for Imaging Science and Technology, pp. 108–113.

Seymour, J., 1995. The why and the how of video-based on-line densitometry. In: *IS&T's Fourth Technical Symposium on Prepress, Proofing, and Printing Proceedings*. Chicago, IL, USA, 8–11 October 1995. Society for Imaging Science and Technology, pp. 23–28.

Sharma, A., 2018. *Understanding color management*. 2nd ed. Chichester: John Wiley & Sons.

Shaw, M., Sharma, G., Bala, R. and Dalal, E.N., 2003. Color printer characterization adjustment for different substrates. *Color Research and Application*, 28(6), pp. 454–467. https://doi.org/10.1002/col.10198.

Simomaa, K., 1987. Are the CCD sensors good enough for print quality monitoring? In: *TAGA 39th Annual Technical Conference Proceedings*. San Diego, CA, USA, 29 March – 1 April 1987. Technical Association of the Graphic Arts, pp. 174–185.

Streckel, B., Steuernagel, B., Falkenhagen, E. and Jung, E., 2003. Objective print quality measurements using a scanner and a digital camera. In: *Proceedings of IS&T's International Conference on Digital Production Printing and Industrial Applications*. Barcelona, Spain, 18–21 May 2003. Society for Imaging Science and Technology, pp. 145–147.

Vapnik, V.N., 1999. *The nature of statistical learning theory*. 2nd ed. Cham: Springer Science & Business Media.

Vapnik, V., Golowich, S.E. and Smola, A.J., 1997. Support vector method for function approximation, regression estimation, and signal processing. In: M. Jordan and T. Petsche, eds. *NIPS'96: Proceedings of the 9th International Conference on Neural Information Processing Systems*. Denver, CO, USA, 3–5 December 1996. Cambridge: MIT Press, pp. 281–278.

Verikas, A. and Bacauskiene, M., 2008. Estimating ink density from colour camera RGB values by the local kernel ridge regression. *Engineering Applications of Artificial Intelligence*, 21(1), pp. 35–42. https://doi.org/10.1016/j.engappai.2006.10.005.

Verikas, A., Bacauskiene, M. and Nilsson, C.-M., 2006. Soft computing for assessing the quality of colour prints. In: A. Moonis and R. Dapoigny, eds. *Advances in Applied Artificial Intelligence: 19th International Conference on Industrial, Engineering and Other Applications of Applied Intelligent Systems*. Annecy, France, 27–30 June 2006. Berlin, Heidelberg: Springer, pp. 701–710. https://doi.org/10.1007/11779568_76.

Watt, P.B., 1956. A densitometer for colour print materials. *The Journal of Photographic Science*, 4(5), pp. 116–120. https://doi.org/10.1080/00223638.1956.11736568.

Xiong, W. and Funt, B., 2006. Estimating illumination chromaticity via support vector regression. *Journal of Imaging Science and Technology*, 50(4), pp. 341–348. https://doi.org/10.2352/j.imagingsci.technol.(2006)50:4(341).

Xuong, N.-h., 1969. An automatic scanning densitometer and its application to x-ray crystallography. *Journal of Physics E: Scientific Instruments*, 2(6): 485. https://doi.org/10.1088/0022-3735/2/6/305.

Yang, B., Chou, H.-Y. and Yang, T.H., 2013. Color reproduction method by support vector regression for color computer vision. *Optik*, 124(22), pp. 5649–5656. https://doi.org/10.1016/j.ijleo.2013.04.036.

Yang, C.-L., Yih, Y., Kuo, Y.-F., Chiu, G. and Allebach, J., 2010. Improving tone prediction in calibration of electrophotographic printers by linear regression: using principal components to account for co-linearity of sensor measurements. *Journal of Imaging Science and Technology*, 54, pp. 50302-1–50302-9. https://doi.org/10.2352/J.ImagingSci.Technol.2010.54.5.050302.

Yazu, Y., Fujihara, M., Takahara, M., Kurata, N., Nakata, A., Yoshimura, H., Ito, T., Fukunaga, M., Kozuki, A. and Tomoi, Y., 2022. Intravascular ultrasound-based decision tree model for the optimal endovascular treatment strategy selection of femoropopliteal artery disease – results from the ONION Study-. *CVIR Endovascular*, 5: 52. https://doi.org/10.1186/s42155-022-00328-9.