Periodicals of Engineering and Natural Sciences ISSN 2303-4521
Vol. 8, No. 3, September 2020, pp.1280-1287

Bootstrap technique for image detection

Alyaa Hashem Mohammed[1], Basim Jamil[2], Fathala Ali[3]
[1,3] Statistics Department, Mustansiriyah University
[2] Computer Science Department, Mustansiriyah University

ABSTRACT

Through more suitability of digital imaging and computer graphics, it suits simpler to convert images content than before without any visually touches for catching these processing issues. Several fake images are created whose content is altered. The research has been reinforced by an application within MATLAB environment of a programmed searching about similar images of the saved image. The research has also been reinforced by a number of forms, pictures, and schemes that clarify the content of the research. The focus of the research lies on two important criteria depending on the content including histogram and statistical criteria of the image for every color. The steps for retrieving process has been clarified starting from statistically analyzing the image and conforming it to the image formed in the database to arrange the images according to their similarity with the target one.

Keywords: Image detection, Bootstrap Technique, Histogram, Statistical criteria of the image

Corresponding Author:

Alyaa Hashem Mohammed
Statistics Department
Collage of administration and Economics
Mustansiriyah University, Iraq
E-mail: aleaa_hashim@uomustansiriyah.edu.iq

1. Introduction

Image or object detection is a computer technique that manipulates an image and detects objects in it. People usually do not distinguish between image detection and image classification, despite the difference is rather clear. If you need to classify image items, we must use classification. On the other hand, if we just need to recognize them, for example, finding out the number of objects in the picture, we should use image detection.
Although digital images are appropriate to use, credibility has many defies due to a lot of fraudulent cases including image forgeries as in fake outputs on human stem-cell research and obtainability of strong photo editing gadgets. Many images touch up methods have become useful, which can be used to produce wonderful artwork [1-3]. Fakes are formed whichever through mixing dual or more photos or changing a raw image. Since image process occurs at the pixel level, detection process is simple as before the digital era. Difficult fakes can be detect using algorithms that notice differences at the bit level [4].
Fake image is becoming a growing threat to information reliability. Through ubiquitous existence of various strong image software tools and smartphone apps such as Photoshop, GIMP, Snapseed, and Pixlr, it has become very easy to manipulate digital images. The field of Digital Image Forensics aims to develop tools that can identify the authenticity of digital images and localize regions in an image which have been tampered with [5]. In this paper, we proposed a new mechanism to detect fake images. The experimental results show that the proposed system achieves accuracy of 100% for fake detection. The experimental tests prove that this proposed method of detection using bootstrap technique is powerful than the classical detection system. Hence, the bootstrap regression model is best standards model for this purpose.

2. Extraction of statistical feature

Characteristics of density grade can be gotten from texture feature in image like flatness, contrast, smoothness, uniformity. Therefore, retrieving image stands for a strong way to reinforce database, and it is useful in similarity calculation of images. Generally, texture feature parameters such as standard deviation, mean, skewness, and kurtosis can be straightforwardly calculated from a color histogram. The last one refers to distribution of colors in image such that the axis (X) signifies the density grade (xi) and the axis (Y) signifies pixels in (xi) [6].

$$p(x_i) = \frac{number\ of\ pixels\ in\ x_i}{total\ number\ of\ pixel\ in\ image}$$

Where prob. refer to probability estimated and $p(x_i)$ $is\ the$ prob xi we can calculated intensity level, as follows: If $p(x_i)$ is the prob. of (x_i) and the L level is total, the mean (μ) is computed density rate value. On the other hand, if (μ) is large, this will lead to brightness of image and this image will be swarthy. μ can be determined by:

$$\mu = \sum_{i=1}^{1} x_i P\,(x_i)$$

The standard deviation (Sd) demonstrates the disparity of gray grade intensity. Low value of the (Sd) demonstrates low variance, while high value demonstrates a high variance as follows:

$$Sd = \sqrt{\sum_{i=1}^{L} (x_i - \mu)^2 P(x_i)}$$

A third feature is a skewness, it demonstrates its density values. Namely, it stands for the measurement of inequality of density grade. Left side of (μ) acts affirmative value, while other side represents passive value. Further, zero amount demonstrates relative eqiuvalent on both sides of (μ). This is characterized by:

$$Skew = \frac{1}{(sd)^3} \sum_{i=1}^{L} (x_i - \mu)^2 P(x_i)$$

Lastly, kurtosis is defined as:

$$g_2 = \frac{m_4}{m_2^2} - 3 = \frac{\frac{1}{n}\sum_{i=1}^{n}(x_i-\mu)^4}{(\frac{1}{n}\sum_{i=1}^{n}(x_i-\mu)^2)^2}$$

Where m_4 is the fourth sample around the (μ), $m_2$is the second sample around the (μ) (sample variance), and x_i is ith value. This formula has simpler representation as follows:

$$g_2 = \frac{1}{n}\sum_{i=1}^{n} z_i^4 - 3$$

Where the zi values are standardized data values using Sd show by apply rather n-1 in the denominator [7].

3. Experimental side

For bootstrap technique for verifying image quality, its examination means recognizing the factual strategies that can be utilized to gauge the nature of conventional (counterfeit) images and their similarity with the first picture. This is done using focal inclination measures (mean, median, mode, skewness and kurtosis) and simple linear regression.

3.1. Simple linear regression

It is a technique uses to examine the impact of quantitative fickle called the responsible fickle [8]. It uses to examine and investigate the impact of a quantitative fickle known as dependent variable. This paradigm can be explained by:

a : regression constant or Part lump of the axis y

b : Straight line slope

The values of X and Y are defined as follows:

$$b = \frac{\sum \frac{xy}{n} - \bar{x}\bar{y}}{s_x^2}$$

$$a = \bar{y} - b\bar{x}$$

3.2 Bootstrap fundamentals

Re-sampling is a statistical method used to determine the quality of data. This method relies on the withdrawal of many samples from the original sample with a return of up to a thousand samples, and a statistical estimation is done with each sample.

Assume that there is a set of observers $(\mathbf{w}_1, \mathbf{w}_2, \dots, \mathbf{w}_\mathbf{n})$, where: $\mathbf{w_i} = (\mathbf{y_i}, \mathbf{x_{ji}})$; then the following steps describes the mechanism of bootstrap:

1- Calculate n bootstrap samples $(w^{(b)}1, w^{(b)}2, \dots, w^{(b)}n)$ and a probability return $\left(\frac{1}{\mathbf{n}}\right)$ for every ($\mathbf{w_i}$).

$$\mathbf{w_i}^{(\mathbf{b})} = \left(\mathbf{y_i}^{(\mathbf{b})}, \mathbf{x_{ij}}^{(\mathbf{b})}\right)$$

$$\mathbf{y_i}^{(\mathbf{b})} = \left(\left(\mathbf{y}_1^{(\mathbf{b})}, \mathbf{y}_2^{(\mathbf{b})}, \dots, \mathbf{y_n}^{(\mathbf{b})}\right)\right)$$

$$x_{ji}^{(b)\prime} = \left(x_{ji}^{(b)}, x_{ji}^{(b)}, \dots, x_{ji}^{(b)}\right)'$$

Where: $i = 1, 2, \dots, n$.

$j = 1, 2, \dots, k$.

2- Calculate the amount of (ols) from the sample bootstrap:

$$\hat{B}^{(b)} = (X^{(b)\prime} X^{(b)})^{-1} X^{(b)\prime} Y^{(b)} \dots \dots \dots \dots (1)$$

3- Repeat steps 1, 2; $r = 1, 2, \dots, B$., where B Represents frequents number.

4- Calculate probability distribution $\mathbf{F}(\hat{\mathbf{B}}^{(\mathbf{b})})$ for bootstrap samples:

$$\hat{\mathbf{B}}^{(b_1)}, \hat{\mathbf{B}}^{(b_2)}, \dots, \hat{\mathbf{B}}^{(b_B)} \qquad \hat{\mathbf{B}}^{(\mathbf{b})} = \sum_{\mathbf{b}=1}^{\mathbf{B}} \frac{\hat{\mathbf{B}}^{(\mathbf{br})}}{\mathbf{B}} = \overline{\hat{\mathbf{B}}}^{(\mathbf{br})}$$

The equation of regression bootstrap will be as:

$$\hat{\mathbf{y}} = \mathbf{x}\hat{\mathbf{B}}^{(\mathbf{b})} + \varepsilon$$

5- Calculate bootstrap standard error:

$$\mathbf{se}\left(\hat{\mathbf{B}}^{\mathbf{b}}\right) = \sqrt{\frac{\sum\left(\hat{\mathbf{B}}^{(\mathbf{br})} - \hat{\mathbf{B}}^{(\mathbf{b})}\right)^2}{\mathbf{B}}} \qquad \hat{\mathbf{B}}^{(\mathbf{b})} = \sum_{\mathbf{b}=1}^{\mathbf{B}} \frac{\hat{\mathbf{B}}^{(\mathbf{br})}}{\mathbf{B}}$$

6- bootstrap estimate of covariance

$$\mathbf{cov}\left(\mathbf{B}_1^{(\mathbf{b})}, \mathbf{B}_2^{(\mathbf{b})}\right) = \frac{1}{\mathbf{B}} \sum \left(\mathbf{B}_1^{(\mathbf{br})} - \hat{\mathbf{B}}^{(\mathbf{b})}\right)\left(\hat{\mathbf{B}}_2^{(\mathbf{br})} - \hat{\mathbf{B}}^{(\mathbf{b})}\right) / \mathbf{B} - 1$$

7- The limits of non-scientific confidence are also called percentile interval which are formed by the bootstrap sample (α/2 %, 1-**α**/ 2 %) [5-6].

$$\hat{\mathbf{B}}^{(\mathbf{br})}_{\mathbf{lower}} \prec \mathbf{B} \prec \hat{\mathbf{B}}^{(\mathbf{br})}_{\mathbf{upper}}$$

Where, $\mathbf{lower} = {}^{\alpha}\!/_{2}\,\mathbf{B}$, $\mathbf{upper} = \left(1 - {}^{\alpha}\!/_{2}\right)\mathbf{B}$

4. Implementation

In this research, the original image was selected and three processing levels were performed as explained by Figure 1. The measure similarity value between the main photo and fake photo was examined based on the following:

1- MATLAB programming language was used for reading the image.
2- Pre-processing for the image is performed, and the resulting image is considered as the traditional image.
3- Extract statistical feature of both original image and traditional image by using central tendency measures (mean, median, mode, skewness and kurtosis).
4- Applying bootstrap-linear simple regression for measuring the similarity degree between both images. Re-sampling estimation according to Monte-Carlo method-based bootstrap includes the following steps:

Assuming that there are data such as: $\mathbf{X} = \{\mathbf{x}_1, \mathbf{x}_2, \ldots\ldots, \mathbf{x}_n\}$.

Where: X represents any statistic. Therefore,

1-The values of those data are entered into the computer.

2-Bootstrap sample is drawn of a size n by drawing a simple random sample of size n with return capability. One sample is chosen each time with return capability according to a random basis as follow:

a) Generate random numbers (r) between 0.000 and 0.999.
b) The random number (r) is multiplied by sample size (n).
c) Close (nr) to the nearest integer to get a single digit number in the sample.
d) The corresponding individual values are selected for this number.
e) Repeat steps (a - d) n - times to allow frequent selection of any value more than once, to get bootstrap sample $\{\mathbf{x}^{\mathbf{b}}{}_1, \mathbf{x}^{\mathbf{b}}{}_2, \ldots\ldots, \mathbf{x}^{\mathbf{b}}{}_{\mathbf{n}}\}$

3- The statistic estimation ($\hat{\mathbf{x}}^{\mathbf{b}}$) is calculated for the bootstrap sample as $\{\hat{\mathbf{x}}^{\mathbf{b}}{}_1, \hat{\mathbf{x}}^{\mathbf{b}}{}_2, \ldots\ldots, \hat{\mathbf{x}}^{\mathbf{b}}{}_{\mathbf{n}}\}$.

4- Steps 2 and 3 are repeated many times (m), then (m) is calculated based on [7].

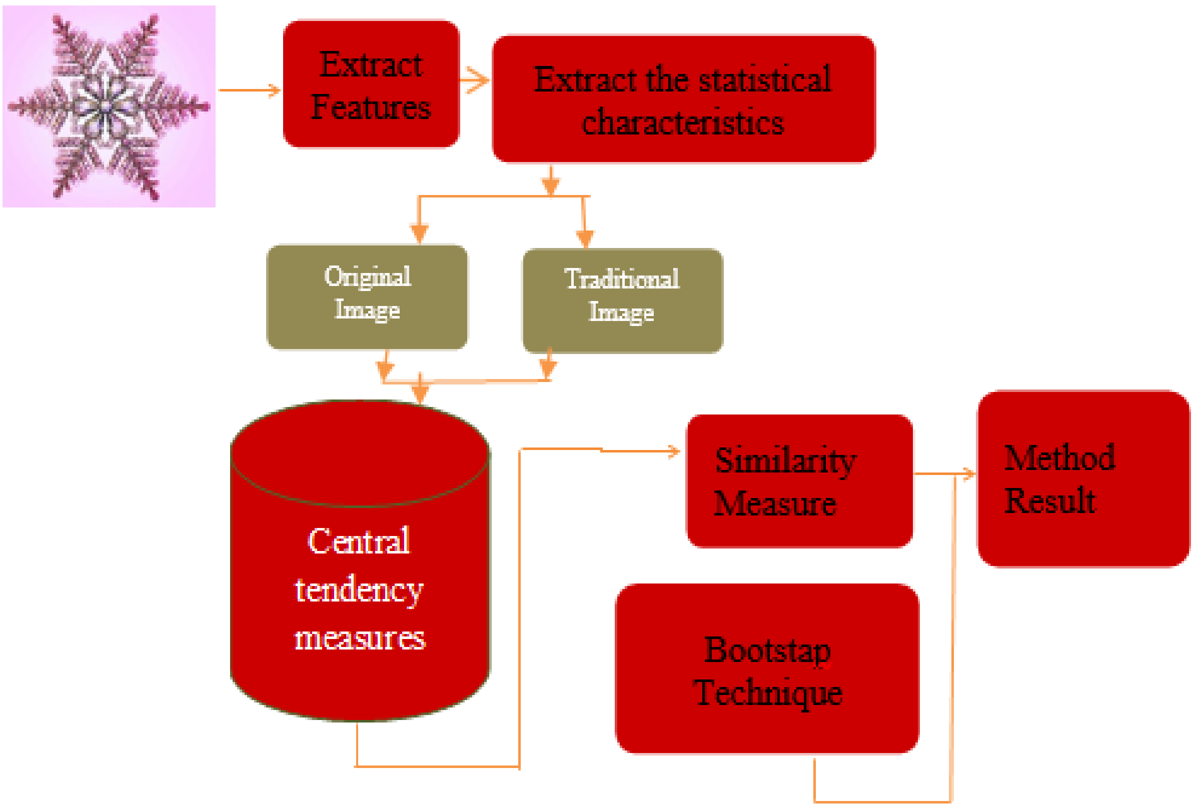
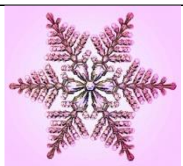
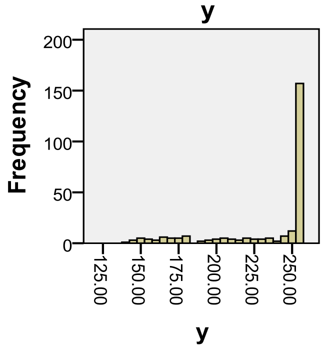
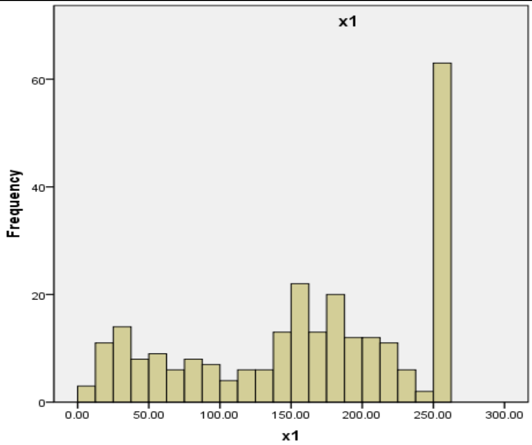
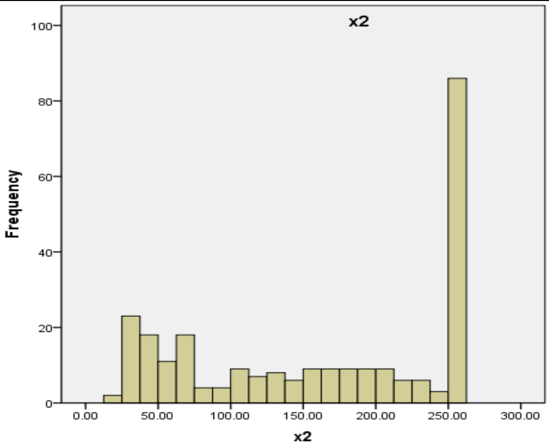
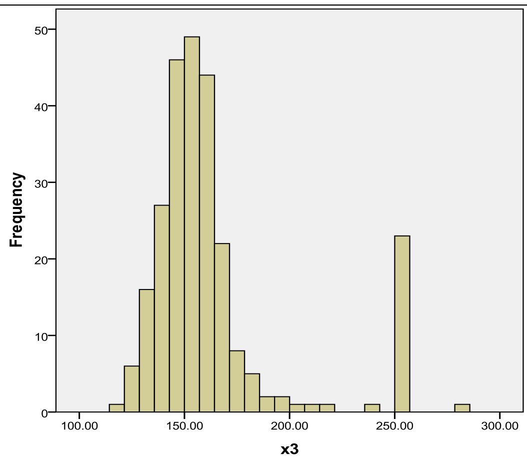


Figure 1. General block diagram of proposed system

Table 1. Histogram and statistical characteristics

	image	Histogram	statistical characteristics
y			Mean 234.3984 Median 255.0000 Mode 255.00 Std. dev 33.65403 Skew -1.438- Kurtosis .593
X1			Mean 161.3281 Median 172.0000 Mode 255.00 Std. dev 79.00649 Skew -.425- Kurtosis -1.044-
X2			Mean 161.2500 Median 176.0000 mode 255.00 Std. dev 86.34141 Skew -.254- Kurtosis -1.536-
X3			Mean 163.6641 Median 155.0000 Mode 255.00 Std. dev 33.64364 Skew 1.928 Kurtosis 2.981

As shown above, the original image Y has the arithmetic mean of (234.3984), the standard deviation of (33.65403), a median of (255.0000), and mode of (255). The image X3 got the highest mean (163.6641) and a standard deviation of (33.64364). The image X2 has got the lowest arithmetic mean of (161.2500) and a standard deviation of (86.34141).

Table 2. Bootstrap statistic

		Bootstrap[a]			
		Bias	Std. Error	95% Confidence Interval	
				Lower	Upper
y	**Mean**	.0970	2.1389	230.2310	238.5818
	Median	.0000	.0000	255.0000	255.0000
	Mode	.0981	2.281	321.1000	301.1000
	Std. Deviation	-.17570-	1.72204	30.07716	36.71966
	Skewness	-.010-	.166	-1.790-	-1.144-
	Kurtosis	.063	.553	-.288-	1.872
X1	**Mean**	.2879	4.9553	151.7900	171.2181
	Median	.1360	6.6043	158.5000	184.5000
	Mode	0.318	7.7140	174.1902	201.387
	Std. Deviation	-.11631-	2.41121	73.88107	83.46056
	Skewness	-.006-	.091	-.609-	-.258-
	Kurtosis	.016	.138	-1.271-	-.738-
X2	**Mean**	.3327	5.4170	150.6862	172.2998
	Median	-.1835-	10.7018	153.5000	194.0000
	Mode	0.402	12.6012	165.2002	200.0001
	Std. Deviation	-.13494-	1.88403	82.38021	89.68001
	Skewness	-.007-	.106	-.479-	-.047-
	Kurtosis	.016	.081	-1.655-	-1.335-
X3	**Mean**	.0612	2.1067	159.5862	168.0257
	Median	-.0870-	1.1891	153.0000	157.0000
	Mode	0.0820	3.0021	200.0110	157.1210
	Std. Deviation	-.09643-	2.41332	28.37970	38.22911
	Skewness	.003	.190	1.560	2.343
	Kurtosis	.093	1.080	1.271	5.655

Table 2 shows the estimation of the coefficients of central tendency measures using the bootstrap method based on the average error squares for each measurement. From Table 2, we noticed the following:

1. The values of (μ), md and mode are unequal. Mode values are larger than the arithmetic mean values.

2. The values of (μ), md and mode are unequal. Mode is greater than the median.
3. The values of (μ), md and mode are unequal. Mode is greater than the median.
4. The mode and the median are equal.
5. The kurtosis curve of (y, x1, x2) is negative while the kurtosis curve x3 is positive.
6. The skewness coefficient of (x3) was (2.9) that is close to the skewness coefficient of the normal distribution, which has a value of (3).
7. The values of the central tendency measurements are between the confidence intervals that are estimated by the bootstrap method, thus indicates that the results are accurate.

5. Impact hypotheses test

This section aims to test the impact hypotheses among the research variables depend on the coefficient (F) and simple regression testing for examining the effect of regression neutralization. This can be realized according to the following hypotheses:

5.1. The hypotheses: (There is a great effect of X-image over Y-original image)

Table 3 below shows that rotating photos has had a great impact on the image as a decision factor (R2) is (0. 393) after the rotation process. This means that its post-processing for the image is interpreted as that the changing in the Y-image is about (39.3%), coefficient (F) was (57.64), and the fixed value (a) has reached (175.852). That is, the amount of the original photo estimate has a fixed value even if there is no tendency factor, and B was (0.363). Namely, altering a single unit of a zooming out procedure will alter the original photo size by (0.289). So, there is no justification for rejecting hypothesis. There is great impact for minimizing procedure of the premier photo. It is according to Table 3:

Table 3. Impact results of x1-image on y- image utilize simple regression bootstrap

Model	Unstandardized Coefficients		Standardized Coefficients		
	B	Std. Error	Beta	t	Sig.
1 (Constant)	175.852	2.513		69.976	.000
x1	.363	.014	.852	25.931	.000

5.2. The hypothesis of a significant effect x2 on original image y

Table 4 displays coefficient of determination (R2) for the zoomed image is (0. 124). This means that there is (12.4%) of variation in original image, the factor (F) is (51.266), fixed value (a) has reached to (72.64). That is, value of original image is estimator based on a fixed value (a) even if there is no tendency factor, and B is (72.64). Altering a single unit of a zooming out procedure will alter the original photo size by (0.289). So, there is no justification for rejecting the assumption. There is a great impact of minimize original photo as shown in Table 4. Unless otherwise noted, bootstrap results are based on 1000 bootstrap samples.

Table 4. Impact results of zooming out of original photo utilizing simple regression

Model			Bootstrap[a]				
						95% Confidence Interval	
		B	Bias	Std. Error	Sig. (2-tailed)	Lower	Upper
1	(Constant)	72.64	-.040	3.698	.001	176.915	192.073
	x2	0.289	.000	.016	.001	.275	.341

5.3. The hypothesis of a significant effect x3 on original image y

Table 5 displays a removal process of background that has realized a significant effect on original image, where R 2 =0.766. Thus, from the interprets, there is a variation of about (76.6%) in the image. In addition, it is found the factor (F= 36.25) and the fixed value (a) has reached to (299.034). That is, value of original image is fixed even if there is no tendency factor, and B is (-0.395), which indicates that altering a single unit of removing the image's background will alter the original photo size by (-0.395). Consequently, there is no justification for rejecting hypothesis (i.e., there is a great effect for removing background from original image). So, it is acknowledged according to these results as in Table 5. Unless otherwise noted, bootstrap results are based on 1000 bootstrap samples

Table 5. Impact results of background mopping from original photo utilize simple regression analysis

Model			Bootstrap[a]				
						95% Confidence Interval	
		B	Bias	Std. Error	Sig. (2-tailed)	Lower	Upper
1	(Constant)	299.034	.263	13.725	.001	274.370	327.235
	x3	-.395	-.002	.090	.001	-.577	-.228

6. Conclusion

The need for efficient detection has increased tremendously in many applications areas such as biomedicine, military, commerce, educations and web image classification. In this paper, we have proposed a new mechanism to detect systems which is mainly based on two procedures. The first procedure relies on extracting the statistical feature of both original, traditional images by using central tendency measures (mean, median, mode, skewness and kurtosis). The second procedure relies on the bootstrap technique -linear simple regression for measuring the similarity degree between both images. The experimental tests prove that this proposed method of detection using bootstrap technique is powerful than the classical detection System.

References

[1] J. I. Lubna, and S. A. K. Chowdhury. "Detecting Fake Image: A Review for Stopping Image Manipulation." In *International Conference on Computational Intelligence, Security and Internet of Things*, pp. 146-159. Springer, Singapore, 2019.

[2] A. Kunbaz, S. Saghir, M. Arar, and E. B. Sönmez. "Fake Image Detection Using DCT and Local Binary Pattern." In *2019 Ninth International Conference on Image Processing Theory, Tools and Applications (IPTA)*, pp. 1-6. IEEE, 2019.

[3] Y. Guo, X. Cao, W. Zhang, and R. Wang. "Fake colorized image detection." *IEEE Transactions on Information Forensics and Security*, Vol.13, No. 8 (2018): 1932-1944.

[4] N. El Abbodi, A. Hassan, M. Al-Nwany, "Blind Fake Image detection", *International Journal of computer Science Issues*, vol.10, Issue 4, No1.,2013.

[5] T. Mohamed, J. Bunk, L. Nataraj, J. Bappy, A. Flenner, B. Manjonath, S. Chandrasekaran, A. Chowdhury, L. Peterson, "Boosting Image Forgery Detection Using Resampling Features and copy-move Analysis", ar Xiv:1802.03154v2[cs.CV], 2018.

[6] A. Al- Achi, "The Student's t- Test: A Brief Description Research & Reviews", *Journal of Hospital and Clinical pharmacy*, Vol.5, No.1, p.1, 2019.

[7] P. Kothyari, S. Dwivedi, and H. L. Mandoria. "Content-Based Image Retrieval Using Statistical Feature and Shape Extraction." *IUP Journal of Information Technology,* Vol. 12, No. 3, p.62, 2016.

[8] J.H. Shahla, T. Ahmed, M.S. Mazen, "Trademark Image Retrieval using Transfer learning", J. Eng. Appl. Sci, Vol.14, No.18, pp.6897- 6905, 2019.