A Novel Method for Forgery Detection on Lung Cancer Images

Bilgehan GÜRÜNLÜ[1] , Serkan ÖZTÜRK[2]

[1]Informatics Department, Kahramanmaras Sutcu Imam University, Kahramanmaras, Turkey
[2] Computer Engineering Department, Erciyes University, Kayseri, Turkey
Corresponding Author: gurunlu@ksu.edu.tr

Research Paper Received: 27.07.2022 Revised: 09.09.2022 Accepted: 15.09.2022

Abstract—With the increase in lung cancer cases in recent years, rapid advances have been made in imaging technologies for lung cancer detection. Thanks to these advances in image processing and medicine, more successful disease diagnosis is achieved. On the other hand, the security of these images is one of the issues that are overlooked or little thought about in this field. The security of images such as Magnetic Resonance Imaging and Computed Tomography is as important as disease detection. Manipulations such as cyber attacks, commit insurance forensic and destruction of evidence can be carried out on health images for various purposes. This problem is included in the study area of both image processing and information security. In this study, we developed a new image forgery detection method based on Center Symmetric Local Binary Pattern texture extraction algorithm, which has not been used on lung cancer images before as far as we known. We tested this method that we have developed on a very up-to-date lung cancer image data set. Although the success of the method is the first study, it is satisfactory. The experimental results of the proposed method show that our method can be used in this field.

Keywords—information security, lung cancer images, computed tomography, passive image security

1. Introduction

As a result of the developments in the area of image and video processing in recent years, successful image and video editing programs have started to be used on multimedia. By using these editing programs, even people who do not have basic knowledge in image or video processing can easily make the changes they want on photos, pictures and videos. Through these editing programs, while end users make changes on photos and videos for entertainment purposes in daily life, professional media organizations (magazines, newspapers, news sites, televisions) use them to make images better or more interesting. On the other hand, malicious people commit different crimes by using such programs on images.

Lung cancer detection studies is one of the areas where technological developments are most adapted and progress is made. Even until the first quarter of the twenty-first century, the technological possibilities of medical oncology and radiation oncology were quite limited. In particular, devices that serve to show the internal structure of the human body were not sufficiently developed. Lung

cancer images, in its simplest form, is to make the internal structure of the human body visible by various methods. The basic devices that fall into this definition can be listed as Radiography Devices, Magnetic Resonance Imaging(MRI) Devices, Nuclear Medicine Imaging Systems, Computed Tomography(CT), and Ultrasound. Today, these devices are produced by a complex synthesis of medical physics, biomedical, electronic engineering and computer science, as well as anatomy and physiology knowledge. Specific protocols and systems are also being developed for archiving and storing lung cancer images [1]. On the other hand, when the studies on disease detection are examined, it is seen that there are also machine learning-based studies in recent years, apart from the algorithms developed based on classical methods [2].

Multidimensional image data in the lung cancer studies field has also increased with progressive developments in the field of image and signal processing in recent years. With these developments, two different problems arise in the field of lung cancer imaging. The first of these problems is that lung cancer images have a more complex structure, both quantitative and content wise, than non-healthcare images. By looking at the whole of the non-healthcare images, oddities on the image can be noticed. In lung cancer images, however, it is difficult for oncologist to see the finding on a small image and diagnose the disease [3].

The second problem on lung cancer images is related to the data security of these images. It is the fact that malevolent people can attack on cancer images and modify them, resulting in oncologist writing false reports and giving people the wrong treatment about their health, leading doctors to decide on the wrong prognosis. While most people edit images for fun, if someone's intention is to harm an innocent patient through the manipulation of lung cancer images, that person may face some real-world consequences detecting whether there is an attack on the lung cancer image and if there is a forged, detecting the region and ensuring the accuracy of the image has been both a security problem and a current study area of image, signal processing science [4]. This negative situation; In the field of computer science, information security, image and video processing, signal processing has created a new field of study that creates a common working area: image forgery detection. As for the detection of such attacks on images, when the methods developed in the literature are examined, it is seen that they are grouped in two different ways. These are active methods and passive methods. In active methods (watermarking, steganography, etc.) preliminary information on the image is needed in advance, while preliminary information is not needed in passive methods. Passive methods are also divided into two sub-studying areas, copy-move forgery and image splicing forgery [5].

The most common image forgery is the copy-move image attack. In copy-move forgery, one or more regions are copied on the same image and pasted to another region on the image. However, in the image splicing forgery, two or more regions are cut through different images and combined with the image to be counterfeited [5]. Studies in the field of copy-move forgery solve forgery problems with two different methodologies: block-based and keypoint-based. Although both methodologies have advantages and disadvantages compared to each other, block-based forgery detection methods are more often used. Since cancer images are important to everyone, in this study, a block-based method was proposed using the Central Symmetric Local Binary Pattern (CSLBP) [6] feature extraction algorithm on lung cancer images in the representation of blocks at the feature extraction stage.

The proposed method has been tested in 10 images on an up-to-date lung cancer image dataset [7]. In the following parts of the study, studies conducted in the literature included in Chapter 2, the proposed method included in Chapter 3, experimental results included in Chapter 4, and results and recommendations included in the last chapter.

2. Related Works

The difference and success of methods in block-based copy-move forgery is based on feature extraction stage. For this purpose, many extraction methods in the literature in the field of signal processing are used for block representation and feature extraction. Some of these are; the Discrete Cosine Transform (DCT), Discrete Wavelet Transform (DWT), Hu moments (HU), principal component analysis (PCA), Kernel-PCA $(KPCA)$, Fourier-Mellin Transform (FMT) and various combinations of these are used for copy-move image forgery [8]. Fridrich [9] brought the first study in this field to the literature using DCT. In Fridrich's study, the image was divided into equal blocks and the DCT of each block was calculated. DCT coefficients and the feature vector of the image were created, and these coefficients were sorted alphabetically to determine similarities between the blocks. In another study [10], proposed an algorithm based on Discrete Wavelet Transform (DWT) for detecting copy move forgery. In this study, after the feature of each block were obtained by DWT by dividing the image into equal blocks. Wang et.al [11] proposed Hu's moments as features extraction and showed robustness for post-processing, including rotation, scaling, and translation in the copy move forgery. Only the first 4 moments are used to reduce computational complexity. Bayram et.al [12] proposed Fourier-Mellin Transform (FMT) to compute 45 block based features and matching stage was reduced using the counting bloom filter vice of lexicographical sorting.

Another study carried out in the literature is the Local Binary Pattern (LBP) feature extraction method, which is often used in texture extraction [13]. The Central Symmetric Local Binary Pattern $(CSLBP)$ algorithm [6], used in the method proposed in this study, which we do on lung cancer images, is an algorithm that works on a neighborhood basis, as in the LBP algorithm. The $CSLBP$ method [6] was developed with inspiration from the LBP pattern extraction operator [13].

The security policies of health systems are lagging behind. While many health images, especially lung cancer images, are protected in terms of data access, data security (availability, integrity) is ignored [14], [15]. Even the leading companies that manufacture imaging devices do not store their lung cancer images encrypted on their servers. In fact, many imaging devices do not support encoding lung cancer images [16]. This situation jeopardizes the safety of lung cancer images. On the other hand, in recent years, studies in the field of information security have also focused on data anonymization [17], [18].

When the research conducted in the field of copy move forgery detection in the literature in recent years are examined, it is seen that there are 2 different methods. These methods are classical methods (DCT, DWT, LBP etc.) and learning based (deep learning, CNN etc.) methods. A large number of images are needed in learning-based methods [19]. In classical methods, even 5 images are sufficient to measure the average performance [20]. Since we proposed a classical method in this study, 10 images were sufficient to determine the success of our algorithm.

3. Proposed Method

Before moving on to the proposed method, explaining the algorithmic logic of the Local Binary Pattern (LBP) algorithm and then explaining the operation of the Central Symmetric Local Binary Pattern $(CSLBP)$ algorithm will provide a better understanding of the proposed method.

The LBP was originally developed as a gray scale for pattern analysis. The LBP operator calculates by comparing the neighborhood values of the central pixel with respect to the 3×3 neighborhood. For each neighborhood, the LBP code is generated for each pixel value in the image. Rule of thumb for LBP code generation; If the center pixel is greater than or equal to the neighboring pixel, the neighboring pixel takes the value '1'. If the center pixel is smaller than the neighboring pixel, the neighboring pixel will take the value '0'. After these pairwise comparisons, starting from the upper left corner of the matrix, the LBP code is generated clockwise, respectively. This code, which is produced by comparison between the center pixel and the neighboring pixel, will be in the range of 0-255 (8 bits). The equation used to calculate the LBP code is given in Equation (1). The values of the $s(t)$ function in this equation are shown by Equation (2). The x value specified in the equation gives the position of the center pixel. x_i represents the position of the pixel with index i. Generating the sample LBP code is shown in Figure 1 below.

$$LBP_{P,R}(X_c, Y_c) = \sum_{p=0}^{p-1} S\left(i_p - i_c\right) 2^p \qquad (1)$$

$$s(x) = \begin{cases} 0, & x < 0 \\ 1, & x \geq 1 \end{cases} \qquad (2)$$

Different neighborhoods of the LBP operator can be calculated according to the 8- and 16-bit distances. After calculating the LBP operator for each pixel in the image, a new LBP image is obtained according to the pixel values varying in the range of 0-255. Then the histogram of the image is taken and a 256-bit texture distribution is found for a particular texture. Although LBP has many successful and widespread applications in the field of image processing, it also brings disadvantages such as the length of the histogram size and being affected by neighboring pixel noises. Therefore, many different feature extraction algorithms have been developed in the literature. $CSLBP$, which was inspired by LBP to remove the histogram length disadvantage of LBP, is a feature extraction algorithm developed to reduce the histogram length. $CSLBP$; It is used in many areas such as texture extraction, face recognition, object recognition. Developed with inspiration from LBP, $CSLBP$ works in a similar structure. As in the LBP method, it compares according to the central pixel, but it calculates according to the neighbor that is circularly opposite to it as a neighborhood. Generating the sample $CSLBP$ code is shown in Figure 2.

The LBP operator produces long histograms. Therefore, it has some difficulties in use. For this reason, $CSLBP$ has made a change in the comparison of neighboring pixels. Instead of comparing each pixel with the central pixel, it uses the method of comparing central symmetric pairs as seen in Figure 2. This symmetrical comparison is performed on the same number of neighbors. For example, for 8 neighborhoods, LBP will calculate 256 different patterns, while $CSLBP$ will calculate only 16 patterns. Therefore, it will have a smaller histogram value. The equation used to calculate the $CSLBP$ code is given in Equation (3). The values of the $s(t)$ function in this equation are shown by Equation (4).

$$CSLP_{R,N|T}(x, y) = \sum_{i=0}^{\left(\frac{N}{2}\right)-1} S\left(n_i - n_{i+\left(\frac{n}{2}\right)}\right) 2^i \quad (3)$$

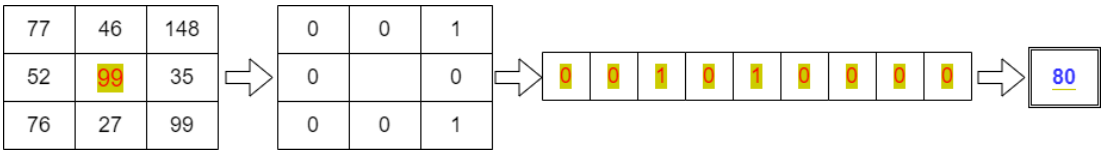

Figure 1. LBP code generation in decimal.

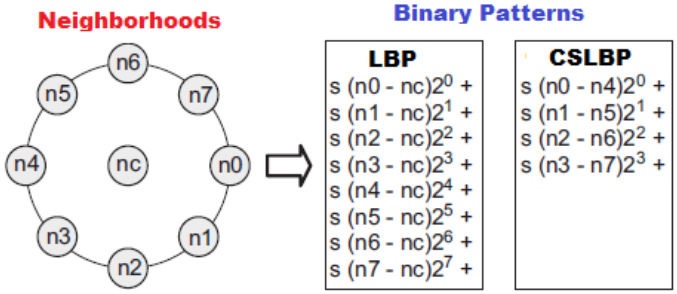

Figure 2. Calculation of LBP and $CSLBP$ for eight-pixel neighborhoods [6].

$$s(x) = \begin{cases} 1, & x > T \\ 0, & 0 \leq 0 \end{cases} \tag{4}$$

Here, n_i and $n_i + (n \backslash 2)$ gray level central symmetric pixels represent pixels with radius R equal to N. It has $N(N-1)+2$ different homogeneous patterns for N neighborhoods. This represents 58 patterns for 8 neighborhoods. Three different metrics are used for CSLBP feature extraction. The radius (R), number of neighboring pixels (N), threshold value (T) value is used for gray level differences. For these metrics in the studies; It is seen that R gives successful results when taken in the range of $\{1, 2\}$, N in $\{6, 8\}$, and T in the range of $\{0, \ldots, 0.02\}$. Feature weighting was not performed for our study.

The success of copy-move forgery detection methods depends on the success of the feature extraction methods used. The simplicity of the models to be used in this area will facilitate important calculations. For this reason, we proposed a method to work on lung cancer images based on the $CSLBP$ algorithm, which has previously achieved successful results in the field of texture recognition, which eliminates the neighboring pixel noise of the LBP.

The pseudocode of our proposed method is shown in Algorithm 1:

Algorithm 1 Algorithm for proposed CSLBP based copy move forgery detection on lung cancer images

Require: $R \times C$ sized image(I), d: block numbers, stopping threshold.

repeat

Step 1. If the $R \times C$ sized image is colored, convert to gray level

$$(Y = 0.299K + 0.5870Y + 0.1140M)$$

Step 2. Divide the $R \times C$ size gray level image into D blocks of $d \times d$ size.

$$0 < i < 1 = (R - d + 1) \text{ and } (C - d + 1)$$

Step 3. Apply the CSLBP to all D blocks of size $d \times d$ and generate a feature vector of length d^2

Step 4. Create | xd^2 dimensional matrix with feature vectors forming rows.

Step 5. Reorder the matrix so that rows of the matrix are ordered from smallest to largest.

Step 6. Calculate shift vector.

// Take the absolute values of two consecutive rows of ordinal matrix and find differences.

Step 7. Calculate the frequency of occurrence of shift vectors and eliminate values less than the threshold values.

Step 8. Calculate the Minkowski distance for remaining values.

Step 9. Eliminate values less than a certain threshold (selected block sizes).

until threshold criteria are satisfied

Resulted image = copied and pasted are marked in forged lung cancer image.

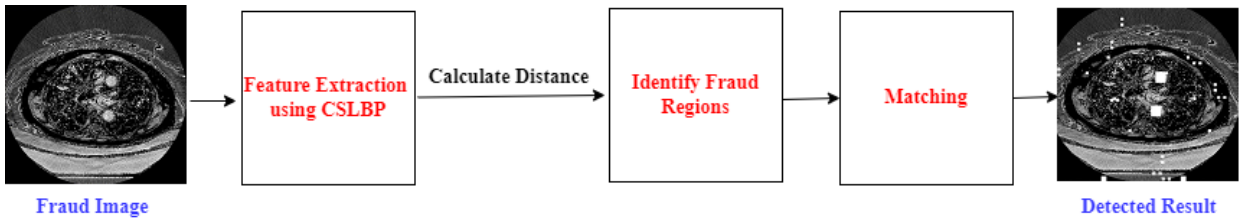

Figure 3. Proposed Method Diagram.

In the next section, we showed the metrics we use in our proposed method and the success comparison according to the known methods.

4. Experimental Results

Conducting experimental studies on lung cancer images and testing the achievements of their developed algorithms is one of the challenges in this area. Because a lot of pre-processing is required on images taken from a real-life hospital environment, these hospital images are replaced by ready-made data sets [21]. With the realization that the safety of lung cancer image data is of vital importance, medical data sets have also begun to be introduced into the literature for researchers in this field. To test the success of the proposed method, we used the random 10 images from the lung cancer image fraud detection dataset [7]. We determined the accuracy of the developed method according to four metrics: TP (True Positive) states that it has been correctly determined that the rate has been tampered with, TN (True Negative) states that it was correctly determined that the rate was not tampered with, FN (False Negative) states that although the forgery was made, it could not be detected, FP (False Positive) states that forgery has been detected although it is not a forgery. The accuracy (ACC) evaluation over these values determines the success of the proposed method and is shown in formula (5) below.

$$ACC = \frac{TP + TN}{TP + FN + TN + FP} \tag{5}$$

The algorithm we suggested, on lung cancer images; For $R = 2$ and $N = 3$, $N = 4$, we got the ACC (Accuracy) values in Table 1 when we tested for 8×8, 12×12 and 16×16 block sizes. When we examine the values in Table 1, it is seen that the accurace rate decreases with the increase of the block sizes, and the accurace rate increases with the increase in the N value.

Table 1.
Accurace Rates of the Proposed Method.

R / N	8x8	12x12	16x16
R=2, N=3 (LBP)	0.7186	0.6883	0.6492
(CSLBP)	0.7209	0.6815	0.6408
R=2, N=4 (LBP)	0.7254	0.6924	0.6547
(CSLBP)	0.7286	0.6879	0.6493

The visual result resulting from the execution of the proposed method is shown in Figure 4.

The first image in the lung cancer detection image in Figure 4 is the original CT image, the second image is that has been frauded CT, and the third image shows copy moved areas detected in blocks. And last image is false positive detected regions. When the visual results are examined, it is seen that FP (False Positive) shows the multiplicity of blocks

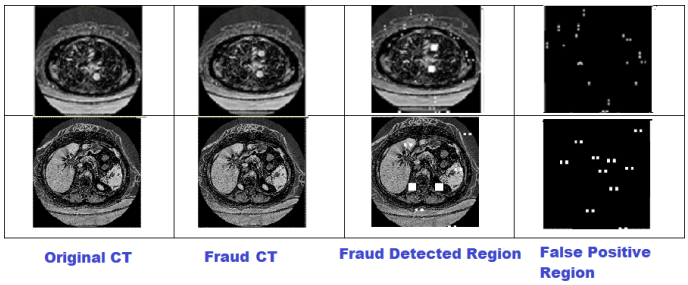

Figure 4. Visual Results of the Method.

that are erroneously marked as forgeries. This, in turn, is a factor that reduces the performance rate.

5. Discussion and Conclusion

Image forgery detection is one of the new and exciting spheres of work in the field of information security. As in many other areas, the security of lung cancer images (MRI and CT) should be ensured against various attacks. Many new algorithms continue to be developed in this area in recent years. Lung cancer images have a very wide field of study, and active research and development work is still ongoing. Standards on this issue are advancing every day.

In this study, a method of detecting lung cancer image forgery was proposed using the Central Symmetric Local Binary Pattern (CSLBP) method, which is a texture-based method in the field of copy-move forgery of lung cancer images. The aim of the study was to minimize the effect of noise on neighboring pixels and to increase the accuracy rate. It has been observed that the performance rates of the CSLBP algorithm may differ according to the selected threshold value. Another notable issue is that if the size of the copied and moved region is less than the size of the selected block, the detection rate decreases. For further studies, different methods (machine learning, deep learning, etc.) can be developed on the data set used. Artificial intelligence optimization studies can also be practiced to find the optimal threshold value for the CSLBP algorithm. While this study was carried out on lung cancer images, the use of this study on real systems in hospitals should also be evaluated among the targets.

Acknowledgments

The authors would like to thank the anonymous reviewers for their useful comments and suggestions.

References

[1] H. Huang, *Pacs-based multimedia imaging informatics: Basic principles and applications*. John Wiley & Sons, 2019.

[2] H. Badem, "Parkinson hastalığının ses sinyalleri üzerinden makine öğrenmesi teknikleri ile tanimlanması," *Niğde Ömer Halisdemir Üniversitesi Mühendislik Bilimleri Dergisi*, vol. 8, no. 2, pp. 630–637, 2019.

[3] M. J. Chuquicusma, S. Hussein, J. Burt, and U. Bagci, "How to fool radiologists with generative adversarial networks? a visual turing test for lung cancer diagnosis," in *2018 IEEE 15th international symposium on biomedical imaging (ISBI 2018)*. IEEE, 2018, pp. 240–244.

[4] T. Seals. (2022) Rsa conference 2019: Ultrasound hacked in two clicks, threatpost. https://threatpost.com/ultrasound-hacked/142601/. Accessed: 29.08.2022.

[5] B. Gurunlu and S. Ozturk, "A survey on photo forgery detection methods," in *ITM Web of Conferences*, vol. 22. EDP Sciences, 2018, p. 01055.

[6] M. Heikkilä, M. Pietikäinen, and C. Schmid, "Description of interest regions with center-symmetric local binary patterns," in *Computer vision, graphics and image processing*. Springer, 2006, pp. 58–69.

[7] Y. Mirsky, T. Mahler, I. Shelef, and Y. Elovici, "{CT-GAN}: Malicious tampering of 3d medical imagery using deep learning," in *28th USENIX Security Symposium (USENIX Security 19)*, 2019, pp. 461–478.

[8] V. Christlein, C. Riess, J. Jordan, C. Riess, and E. Angelopoulou, "An evaluation of popular copy-move forgery detection approaches," *IEEE Transactions on information forensics and security*, vol. 7, no. 6, pp. 1841–1854, 2012.

[9] A. J. Fridrich, B. D. Soukal, and A. J. Lukáš, "Detection of copy-move forgery in digital images," in *in Proceedings of Digital Forensic Research Workshop*. Citeseer, 2003.

[10] P. Yadav and Y. Rathore, "Detection of copy-move forgery of images using discrete wavelet transform," *International Journal on Computer Science and Engineering*, vol. 4, no. 4, p. 565, 2012.

[11] J. Wang, G. Liu, Z. Zhang, Y. Dai, and Z. Wang, "Fast and robust forensics for image region-duplication forgery," *Acta Automatica Sinica*, vol. 35, no. 12, pp. 1488–1495, 2009.

[12] S. Bayram, H. T. Sencar, and N. Memon, "An efficient and robust method for detecting copy-move forgery," in *2009 IEEE International Conference on Acoustics, Speech and Signal Processing*. IEEE, 2009, pp. 1053–1056.

[13] T. Ojala, M. Pietikäinen, and D. Harwood, "A comparative study of texture measures with classification based on featured distributions," *Pattern recognition*, vol. 29, no. 1, pp. 51–59, 1996.

[14] L. Coventry and D. Branley, "Cybersecurity in healthcare: A narrative review of trends, threats and ways forward," *Maturitas*, vol. 113, pp. 48–52, 2018.

[15] M. S. Jalali and J. P. Kaiser, "Cybersecurity in hospitals: a systematic, organizational perspective," *Journal of medical Internet research*, vol. 20, no. 5, p. e10059, 2018.

[16] K. Zetter. (2022) Hospital viruses: Fake cancerous nodes in ct scans, created by malware, trick radiologists. https://www.washingtonpost.com/technology/2019/04/03/hospital-viruses-fake-cancerousnodes-ct-scans-created. Accessed: 30.08.2022.

[17] Y. Canbay, S. Sagiroglu, and Y. Vural, "A new utility-aware anonymization model for privacy preserving data publishing," *Concurrency and Computation: Practice and Experience*, vol. 34, no. 10, p. e6808, 2022.

[18] Y. Canbay, A. İsmetğlu, and P. Canbay, "Covid-19 hastalığının teşhisinde derin öğrenme ve veri mahremiyeti," *Mühendislik Bilimleri ve Tasarım Dergisi*, vol. 9, no. 2, pp. 701–715, 2021.

[19] M. A. Elaskily, H. A. Elnemr, A. Sedik, M. M. Dessouky, G. M. El Banby, O. A. Elshakankiry, A. A. Khalaf, H. K. Aslan, O. S. Faragallah, A. El-Samie *et al.*, "A novel deep learning framework for copy-moveforgery detection in images," *Multimedia Tools and Applications*, vol. 79, no. 27, pp. 19 167–19 192, 2020.

[20] A. Kuznetsov and V. Myasnikov, "A new copy-move forgery detection algorithm using image preprocessing procedure," *Procedia engineering*, vol. 201, pp. 436–444, 2017.

[21] Open-Acess Medical Image Repositories. (2022) Open access. https://www.aylward.org/notes/open-access-medical-image-repositories. Accessed: 15.05.2022.