rmance_for_Leukemia_Cell_Images_using_Support_Vector_Machine.pdf

by

Submission date: 01-Apr-2023 01:42PM (UTC+0700)
Submission ID: 2052750957
File name: rmance_for_Leukemia_Cell_Images_using_Support_Vector_Machine.pdf (263.43K)
Word count: 3678
Character count: 19616

2021 1st International Conference on Electronic and Electrical Engineering and Intelligent System (ICE3IS)

Comparison of Texture and Shape Features Performance for Leukemia Cell Images using Support Vector Machine

1st Yessi Jusman
Department of Electrical Engineering
Faculty of Engineering, Universitas Muhammadiyah Yogyakarta
Yogyakarta, Indonesia
*Corresponding Email:
yjusman@umy.ac.id

2nd Ega Samudra
Department of Electrical Engineering
Faculty of Engineering, Universitas Muhammadiyah Yogyakarta
Yogyakarta, Indonesia

3rd Slamet Riyadi
Department of Informatic Tecnology
Faculty of Engineering, Universitas Muhammadiyah Yogykarta
Yogyakarta, Indonesia

4th Siti Nurul Aqmariah Mohd Kanafiah
School of Mechatronics Engineering, Universiti Malaysia
Perlis, Malaysia

5th Amir Faisal
Department Biomedical Engineering, Institut Teknologi Sumatera
Lampung, Indonesia

6th Rosline Hassan[a], 7th Zeehaida Mohamed[b]
[a]*Department of Haematology*
[b]*Department of Microbiology and Parasitology, Universiti Sains Malaysia*
Kelantan, Malaysia

2021 1st International Conference on Electronic and Electrical Engineering and Intelligent System (ICE3IS) | 978-1-6654-0546-1/21/$31.00 ©2021 IEEE | DOI: 10.1109/ICE3IS54102.2021.9649683

***Abstract*—Leukemia or often called blood cancer is one type of cancer caused by excessive white blood cells. Excessive white blood cells will cause disruption of normal function of other blood cells. To find out leukemia, we can do a physical examination in the form of a blood sample or can also use a spinal cord biopsy. In general, doctors take blood samples to see and look for abnormalities of the white blood cell count. To reduce human error in diagnosing leukemia, the study created two systems that can classify leukemia using the Hu moment invariant (HMI) and Support Vector Machine (SVM) methods and the Grey Level Co-occurance Matrix (GLCM) and SVM methods. Classification system are used to classify acute and normal leukemia image classes using 10-fold cross validation in the sharing of its image data. The best classification results are the GLCM-SVM system with an accuracy value of 99% and the HMI-SVM system produces an accuracy value of 90%.**

***Keywords*—Hu Moments, K-Nearest Neighbor, Leukemia, Acute Leukemia, Normal Leukemia, Support Vector Machine**

I. INTRODUCTION

Leukemia is a cancer of the blood cells originating from the bone marrow. The proliferation of white blood cells usually characterizes it by manifesting abnormal cells in the peripheral blood (blast cells) in excess and causing the suppression of normal blood cells, resulting in impaired function.

Identification and classification of leukemia cancer are crucial because treatment varies according to the subtype of leukemia. The conventional approach to classify cancer based on morphological characteristics has been found to be inadequate due to the underlying complexity and ambiguity in cancer classification. Thus, it takes highly skilled resources to detect differences among growth cells. This procedure is immensely time-consuming and expensive. In other words, such a handling procedure is an inappropriate solution. Cells can appear morphologically the same but react in stark contrast to drugs and cytotoxic treatments.

Several studies have attempted to develop a computer-assisted system with digital image processing methods and different classification methods to help deal with this leukemia problem. Starting from research using hybrid hierarchical classifiers and Fuzzy C Means based on morphological contour segmentation to the application of the watershed algorithm and Gray Level Co-occurrence Matrix (GLCM) in leukemia cells images are discussed [1], [2], [3], [4], [5], [6], [7], [8], [9], [10].

Machine learning system is mostly be applied to classification purpose of biomedical images. The system can be aided the expert to obstacle the time consuming procedure in the manual diagnosis. SVM and K-NN methods are mostly implemented in many research related to classification system for diagnosis purposes [11], [12]. Several machine learning for leukemia have been presented and discussed [13], [14], [15], [16], [17].

The shape of the imagery features has the most crucial role in classifying the images effectively and efficiently. Some studies have used Hu moment invariant (HMI) algorithms in recognizing the shape [18], [19], [20], [21], [22], [23]. Based on the literature review, the implementations of hu moment invariant algorithm for feature extraction are limited. Thus, this research will take the gap in this research topic.

This study discusses the classification of normal white blood cells and acute leukemia cells using the HMI algorithm for feature extraction with the linear type Support Vector Machine (SVM) classification methods namely HMI-SVM methods by comparing the results of the GLCM-SVM methods. The best value of the results can differentiate between acute and normal leukemia cells.

This research is expected to help the medical problems in handling the diagnosis of white blood cell cancer quickly and accurately. Thus, it inhibits white blood cell cancer from spreading to other tissues and eases medical personnel to diagnose the type of leukemia patients suffer, thereby reducing the mortality rate due to delays in handling and diagnosing the type of leukemia.

II. METHODOLOGY

A. System Design

System design is one of the essential stages in applying a research concept. Thus both programs and program results can run as intended. In this study, the system design was carried out using algorithms, and training was performed on research data.

This study required software and hardware as performance support devices to assist in the system design process. The hardware used is demonstrated in Table 1, while the software was MATLAB R2019b using an application on the MATLAB system in the form of a classification learner.

TABLE I. HARDWARE SPECIFICATIONS

Specification	Description
Processor	Intel(R) Core(TM) i5-4200U @ 2,30 GHz
System Type	64-bit Operating System
RAM	8.00 GB

This study utilized 200 images of leukemia cells (Fig.1) consisting of 100 acute images and 100 normal images verified from hospital patients at the Universiti Sains Malaysia as the data. The images were divided into ten datasets to facilitate the training and testing stages. The images were divided into 10-fold using the K-fold cross-validation method; each fold encompassed training image data and testing image data with a ratio of 90:10 comprising 180 training images and 20 testing images with bmp format image.

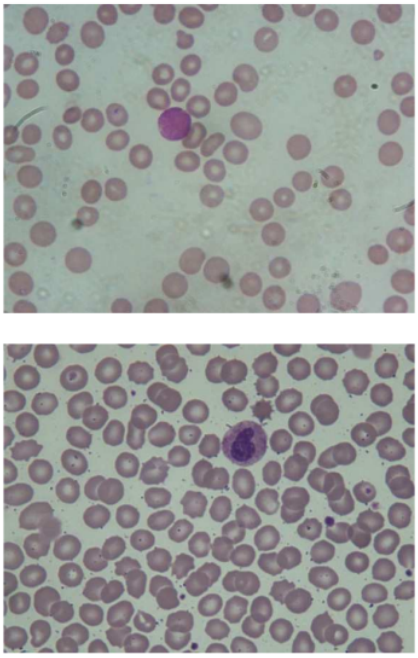

Fig. 1. Example of Leukemia Images

The system design process consisted of two main stages: the training and the testing. The first stage was training on the data training, comprising the feature extraction process with the HMI, and the training process using the linear type SVM. In the SVM training process, a database was obtained in the forms of accuracy, confusion matrix, and training time used as a classification process in the testing stage and comparing the two methods. The stages of the training process are displayed in Figure 2.

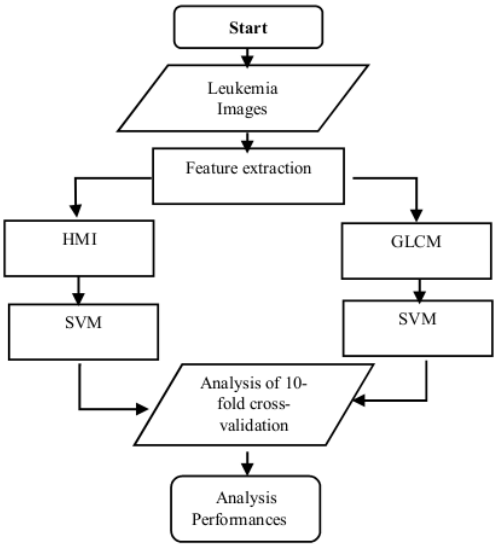

Fig. 2. Design of the training process system

At the testing stage on the testing data, several processes were carried out to classify the images. The testing processes comprised a feature extraction process using the HMI algorithm. The HMI are normally extracted from the outline of the cells in an image. It is used in this research due to the different of the cells shape between normal and abnormal cells based on the shape feature vectors.

The results from the linear type SVM training process was employed in this classification testing process. Figure 3 illustrates the system design flow of the testing process. The detail information of the system is presented.

1. Feature extraction with HMI method

Feature extraction aims to obtain significant data (i.e. shape features vectors) on the images. Hence, cells in the images can be distinguished from one another. The shape features of the cells have the most crucial role in identifying the shape effectively and efficiently to distinguish the images of acute and normal leukemia cells.

The data from the feature extraction of the HMI were in the form of seven feature moments (Hu's seven-moment invariants) in phi values totaling seven values consisting of phi 1, phi 2, phi 3, phi 4, phi 5, phi 6, and phi 7. The seven phi values were analyzed by calculating the average value of each 360 acute and 360 normal images and the standard

deviation value. Image features at the training stage from the feature extraction results were labeled on each image data following its class.

TABLE II. AVERAGE VALUE RESULTS AND S.DEVIATION OF HU MOMENT INVARIANT FEATURES

Features	Classes of Images	
	Acute Leukemia	Normal
Phi 1	0,002 ± 7,164	0,0023 ± 8,973
Phi 2	3,824 ± 2,454	4,4618 ± 3,294
Phi 3	7,869 ± 9,796	7,4790 ± 8,346
Phi 4	7,684 ± 9,322	5,9426± 8,285
Phi 5	1,446 ± 3,567	1,0287 ± 2,610
Phi 6	1,474 ± 1,878	1,236± 1,740
Phi 7	4,488 ± 2,669	-4,377± 1,390

2. Features extraction with GLCM method

Texture are the features are presented in contrast, correlation, energy, and homogeneity features as tabulated in Tables 3-6. The extraction of features is carried out using the Gray Level Co-occurance Matrix (GLCM) method with data input of 200 leukemia images consisting of 2 classes, namely acute leukemia and normal. The extraction of features using GLCM produces data output from 4 features consisting of contrast, correlation, energy, and homogeneity using angles of 0°, 45°,90°, and 135° which will be used as references at the classification stage of the SVM method. The detection system for leukemia with the GLCM-SVM method uses distance of 8 pixels while the quantization level uses 50. Data results from feature extraction obtained for both acute and normal class types.

TABLE III. AVERAGE VALUE RESULTS AND S. DEVIATION OF GLCM FEATURE ANGLE 0°

Features	Classes of Images	
	Acute Leukemia	Normal
Contrast	1,62 ± 0,73	1,70 ± 0,25
Correlation	0,072 ± 0,161	-0,054, ± 0,033
Energy	0,191 ± 0,028	0,132 ± 0,0102
Homogeneity	0,684 ± 0,029	0,613± 0,017

TABLE IV. AVERAGE VALUE RESULTS AND S. DEVIATION OF GLCM FEATURE ANGLE 45°

Features	Classes of Images	
	Acute Leukemia	Normal
Contrast	1,82 ± 1,160	1,60 ± 0,240
Correlation	0,041 ± 0,777	0,021 ± 0,030
Energy	0,186 ± 0,029	0,131 ± 0,011
Homogeneity	0,672 ± 0,031	0,635± 0,017

TABLE V. AVERAGE VALUE RESULTS AND S. DEVIATION GLCM FEATURE ANGLE 90°

Features	Classes of Images	
	Acute Leukemia	Normal
Contrast	1,70 ± 0,87	1,63 ± 0,250
Correlation	0,076 ± 0,132	-0,014 ± 0,033
Energy	0,191 ± 0,029	1,131 ± 0,010
Homogeneity	0,682 ± 0,027	0,624± 0,015

TABLE VI. AVERAGE VALUE RESULTS AND S. DEVIATION GLCM FEATURE ANGLE 135°

Features	Classes of Images	
	Acute Leukemia	Normal
Contrast	1,82 ± 1,140	1,58 ± 0,230
Correlation	0,037 ± 0,838	0,018 ± 0,025
Energy	0,186 ± 0,029	0,131 ± 0,011
Homogeneity	0,671 ± 0,031	0,635± 0,015

3. SVM training data

At the training stage, features as input of the training data were entered into the application on the MATLAB system in the form of a classification learner using the K-fold cross-validation with the linear type SVM, resulting in the value of training accuracy, confusion matrix analysis results, and training time.

4. Data analysis

Based on the training data results using the data preparation method of 10-fold cross-validation, the best weight from each training process, both with the SVM models, were taken to be used as the weight of the system being built. Testing was carried out to determine whether the system built could work and function properly. Data from classification results were in the form of the system's success in distinguishing acute and normal image cells in leukemia regarding the accuracy, specificity, and sensitivity.

5. System testing using testing data

Figure 2 exhibits the system testing process. In the testing data, the identification process was run by feature extraction with the HMI to obtain the 7 phi value and the GLCM to obtain 16 features, with the aim of the system being able to classify the types of leukemia by studying the training data and then classifying the testing data based on HMI and GLCM using the SVM method. The performances of HMI and GLCM features data are compared by using SVM method.

6. Classification using SVM

The classification stage was the decision stage in determining the image class. This study employed two classes, normal and acute. To classify images after

determining the class, the training data were inputted with the testing image data. The classification process utilized the results of the SVM training and the value of the feature matrix from the feature extraction results.

B. Analysis

The success of the classification learner's performance in the classification system using the SVM method was when the system could identify the type of leukemia cells appropriately. The results of the recorded system classification were based on the confusion matrix. Based on the confusion matrix obtained with 10-fold cross-validation the performance of the system was assessed based on the values of accuracy, sensitivity, specificity, and training time of the classification system.

III. Results and Discussions

The testing process was performed to determine the performance of the system design. The testing process results using the classification learner application in the MATLAB program obtained results in accuracy values, and training time. Table 7 demonstrates the training results of HMI-SVM and GLCM-SVM algorithms.

TABLE VII. The Training Results of HMI-SVM and GLCM-SVM

Data Set-N	Classification results of HMI-SVM		Classification results of GLCM-SVM	
	Accuracy	Time	Accuracy	Time
Data Set 1	90.6 %	1.01s	98.3 %	1.7s
Data Set 2	90.6 %	1.05s	98.3 %	1.3s
Data Set 3	91.7 %	1.04s	99.4 %	1.03s
Data Set 4	91.7 %	1.06s	98.8 %	0.8s
Data Set 5	88.9 %	1.07s	99.4 %	1.05s
Data Set 6	88.3 %	1.07s	97.8 %	0.9s
Data Set 7	90.6 %	1.5s	98.9 %	0.9s
Data Set 8	88.9 %	1.3s	98.9 %	0.9s
Data Set 9	91.1 %	1.3s	98.3 %	0.9s
Data Set 10	91.1 %	1.3s	98.9 %	1.0s
Averages	90.35 %	1.17s	98.7 %	1.448s

Based on Table 7, the results of accuracy using the HMI-SVM method resulted in an average accuracy value of 90.35% and required data execution time of 1.17s. Meanwhile, the results of accuracy using the GLCM-SVM extraction method resulted in an average accuracy value of 98.7% and required data execution time of 1,448s.

This stage is done to help the system in learning more detail about the characteristics of the image so that it can distinguish two classes of images. Thus, system testing will result in a high accuracy value. The more images the system learns, the greater the weight of the results that will be produced.

TABLE VIII. Comparison of Testing Results of the HMI-SVM and GLCM-SVM

Data Set-N	HMI-SVM Performances	GLCM-SVM Performances
	Accuracy (%)	Accuracy (%)
Data Set 1	100	100
Data Set 2	90	100
Data Set 3	90	95
Data Set 4	80	100
Data Set 5	100	95
Data Set 6	80	100
Data Set 7	95	100
Data Set 8	100	100
Data Set 9	95	100
Data Set 10	65	100
Averages	90	99

Based on Table 8, SVM classification results of testing data can be seen the SVM using the HMI extraction method produces a variety of accuracy results. In data set1, data set5 and data set 8 produce the highest accuracy value with an accuracy result of 100%. While on the data set 10 produces the lowest accuracy value which results in an accuracy value of 65%. Although there are datasets with the lowest accuracy values of 65%, there are some datasets that have accuracy values above 80% that can help increase the average results in this classification. From the results of the accuracy of Table 8 resulted in an average value of 90% which indicates that using the SVM classification produces a fairly good gloom. From the results of 2 systems SVM using features extraction with GLCM and HMI methods it can be concluded that the best results are obtained using GLCM-SVM.

Based on Tables 7 and 8, it can be seen that the comparison between SVM using the GLCM extraction method with SVM using the HMI feature extraction method has a percentage value with a significant difference despite using the same data GLCM-SVM system, the average value generated by SVM using the HMI method have a lower average value. Accordingly, based on the results of both tests that have been done, it can be concluded that SVM using the GLCM feature extraction method is better when compared to SVM which uses the HMI feature extraction method. However, both systems are able to classify leukemia and work well.

IV. Conclusions

Using Hu Moment Invariant (HMI) or Grey Level Co-occurance Matrix (GLCM) and classification method using the Support Vector Machine (SVM) method is able to classify with an average accuracy value of > 90% of both systems. System design using the HMI feature extraction method can perform feature extraction with 7 features as a reference classification stage. Meanwhile, the system using the GLCM feature extraction method can perform feature extraction with 16 features in the form of contrast, correlation, energy and homogeneity using an angle of 0°, 45°, 90°, 135°. The GLCM-SVM method produces the highest accuracy value of 100%, while the lowest accuracy value is obtained by 95%. For the average value of all datasets, the leukemia classification obtained an accuracy of 99%. HMI-SVM methods produces

the highest accuracy value of 100%, while the lowest accuracy value is obtained by 65%. Accuracy averages of leukemia classification from the overall data set obtained accuracy of 90%. The GLCM-SVM method produces a higher accuracy average value of 99% when compared to the HMI-SVM method which produces an average accuracy value of 90%.

ACKNOWLEDGMENT

This research is supported by Universitas Muhammadiyah Yogyakarta and a research project grant from the Ministry of Research and Technology of the Republic of Indonesia.

REFERENCES

[1] Jabar, F.H., et al., Image Segmentation using A Hybrid Clustering Technique and Mean Shift for Automated Detection Acute Leukaemia Blood Cells Images. Journal of Theoretical & Applied Information Technology, 2015. 76(1).

[2] Putzu, L., G. Caocci, and C. Di Ruberto, Leucocyte classification for leukaemia detection using image processing techniques. Artificial Intelligence in Medicine, 2014. 62(3): p. 179-191.

[3] Neoh, S.C., et al., An intelligent decision support system for leukaemia diagnosis using microscopic blood images. Scientific reports, 2015. 5: p. 14938.

[4] Viswanathan, P., Fuzzy C Means Detection of Leukemia Based on Morphological Contour Segmentation. Procedia Computer Science, 2015. 58: p. 84-90.

[5] Mishra, S., et al. Microscopic image classification using DCT for the detection of acute lymphoblastic leukemia (ALL). in Proceedings of International Conference on Computer Vision and Image Processing. 2017. Springer.

[6] Rawat, J., et al., Classification of acute lymphoblastic leukaemia using hybrid hierarchical classifiers. Multimedia Tools and Applications, 2017. 76(18): p. 19057-19085.

[7] Wang, Q., et al., Spectral-spatial feature-based neural network method for acute lymphoblastic leukemia cell identification via microscopic hyperspectral imaging technology. Biomedical optics express, 2017. 8(6): p. 3017-3028.

[8] Abdeldaim, A.M., et al., Computer-Aided Acute Lymphoblastic Leukemia Diagnosis System Based on Image Analysis, in Advances in Soft Computing and Machine Learning in Image Processing, A.E. Hassanien and D.A. Oliva, Editors. 2018, Springer International Publishing: Cham. p. 131-147.

[9] Negm, A.S., O.A. Hassan, and A.H. Kandil, A decision support system for Acute Leukaemia classification based on digital microscopic images. Alexandria Engineering Journal, 2018. 57(4): p. 2319-2332.

[10] Jusman, Y., et al. Application of Watershed Algorithm and Gray Level Co-Occurrence Matrix in Leukemia Cells Images. in 2020 3rd International Conference on Mechanical, Electronics, Computer, and Industrial Technology (MECnIT). 2020.

[11] Chandra, M.A. and S.S. Bedi, Survey on SVM and their application in image classification. International Journal of Information Technology, 2018.

[12] Deng, Z., et al., Efficient kNN classification algorithm for big data. Neurocomputing, 2016. 195: p. 143-148.

[13] Rawat, J., et al., Computer aided diagnostic system for detection of leukemia using microscopic images. Procedia Computer Science, 2015. 70: p. 748-756.

[14] MoradiAmin, M., et al., Computer aided detection and classification of acute lymphoblastic leukemia cell subtypes based on microscopic image analysis. Microscopy research and technique, 2016. 79(10): p. 908-916.

[15] Srisukkham, W., et al., Intelligent leukaemia diagnosis with bare-bones PSO based feature optimization. Applied Soft Computing, 2017. 56: p. 405-419.

[16] Jagadev, P. and H. Virani. Detection of leukemia and its types using image processing and machine learning. in 2017 International Conference on Trends in Electronics and Informatics (ICEI). 2017. IEEE.

[17] Pan, L., et al., Machine learning applications for prediction of relapse in childhood acute lymphoblastic leukemia. Scientific reports, 2017. 7(1): p. 1-9.

[18] Liu, Y., Y. Yin, and S. Zhang. Hand Gesture Recognition Based on HU Moments in Interaction of Virtual Reality. in 2012 4th International Conference on Intelligent Human-Machine Systems and Cybernetics. 2012.

[19] Zhang, Y., et al., Pathological brain detection based on wavelet entropy and Hu moment invariants. Bio-Medical Materials and Engineering, 2015. 26: p. S1283-S1290.

[20] Sari, B.P. and Y. Jusman. Classification System for Cervical Cell Images based on Hu Moment Invariants Methods and Support Vector Machine. in 2021 International Conference on Intelligent Technologies (CONIT). 2021.

[21] Zhang, Y., et al., Pathological brain detection in MRI scanning via Hu moment invariants and machine learning. Journal of Experimental & Theoretical Artificial Intelligence, 2017. 29(2): p. 299-312.

[22] Zhang, Y.-D., et al., Alcoholism detection by medical robots based on Hu moment invariants and predator–prey adaptive-inertia chaotic particle swarm optimization. Computers & Electrical Engineering, 2017. 63: p. 126-138.

[23] Salleh, M.A.M., S.N.A.M. Kanafiah, and Y. Jusman. Features Extraction to Differentiate of Spinal Curvature Types using Hue

rmance_for_Leukemia_Cell_Images_using_Support_Vector_M...

ORIGINALITY REPORT

11%	8%	8%	4%
SIMILARITY INDEX	INTERNET SOURCES	PUBLICATIONS	STUDENT PAPERS

PRIMARY SOURCES

1	journal2.uad.ac.id Internet Source	2%
2	Syahputra, Ramadoni, and Indah Soesanti. "Power System Stabilizer model based on Fuzzy-PSO for improving power system stability", 2015 International Conference on Advanced Mechatronics Intelligent Manufacture and Industrial Automation (ICAMIMIA), 2015. Publication	1%
3	Chin-Ling Lee, Cheng-Jian Lin. "Integrated Computer Vision and Type-2 Fuzzy CMAC Model for Classifying Pilling of Knitted Fabric", Electronics, 2018 Publication	1%
4	Submitted to National Tsing Hua University Student Paper	1%
5	repository.um.edu.mo Internet Source	1%
6	www.hindawi.com Internet Source	1%

7	dokumen.pub Internet Source	1 %
8	www.angelfire.com Internet Source	<1 %
9	ieeexplore.ieee.org Internet Source	<1 %
10	repository.unikal.ac.id Internet Source	<1 %
11	ojs.kalbis.ac.id Internet Source	<1 %
12	S. Patil Babaso, S.K. Mishra, Aparna Junnarkar. "Leukemia Diagnosis Based on Machine Learning Algorithms", 2020 IEEE International Conference for Innovation in Technology (INOCON), 2020 Publication	<1 %
13	dosen.unimma.ac.id Internet Source	<1 %
14	www.mdpi.com Internet Source	<1 %
15	www.semanticscholar.org Internet Source	<1 %
16	ijettjournal.org Internet Source	<1 %

17	"Advances in Soft Computing and Machine Learning in Image Processing", Springer Science and Business Media LLC, 2018 Publication	<1 %
18	"Multimedia Technology and Enhanced Learning", Springer Science and Business Media LLC, 2020 Publication	<1 %
19	"Prognostic Models in Healthcare: AI and Statistical Approaches", Springer Science and Business Media LLC, 2022 Publication	<1 %
20	A Jamal, S G Putri, A N N Chamim, R Syahputra. "Harmonic Reduction Using Shunt Active Filter for Electrical Installation", Journal of Physics: Conference Series, 2020 Publication	<1 %
21	Arif Muntasa, Muhammad Yusuf. "Three Channels for Gray Level Co-occurrence Matrix (GLCM) to detect Acute Lymphoblastic Leukemia (ALL) Images", 2020 the 3rd International Conference on Control and Computer Vision, 2020 Publication	<1 %
22	Parastoo Taghikhani. "Hybrid Beamforming Transmitter Modeling for Millimeter-Wave	<1 %

MIMO Applications", Institute of Electrical and Electronics Engineers (IEEE), 2020
Publication

23	Roopa B. Hegde, Keerthana Prasad, Harishchandra Hebbar, I Sandhya. "Peripheral blood smear analysis using image processing approach for diagnostic purposes: A review", Biocybernetics and Biomedical Engineering, 2018 Publication	<1 %

Exclude quotes	On	Exclude matches	Off
Exclude bibliography	On		