International Journal of Electrical and Computer Engineering (IJECE)
Vol. 12, No. 4, August 2022, pp. 3692~3699
ISSN: 2088-8708, DOI: 10.11591/ijece.v12i4.pp3692-3699

Acute lymphoblastic leukemia detection approach from peripheral blood smear using color threshold and morphological techniques

Abdullah Al Mamun[1], Md. Jakir Hossen[1], Anik Tahabilder[2], Ahmmad Musha[3], Rehnuma Hasnat[3], Sohag Kumar Saha[4]
[1]Faculty of Engineering and Technology, Multimedia University, Melaka, Malaysia
[2]Department of Computer Science, Wayne State University, Detroit, United States
[3]Department of Electrical and Electronic Engineering, Pabna University of Science and Technology, Pabna, Bangladesh
[4]Electrical and Computer Engineering, Tennessee Technological University, Cookeville, United States

Article Info

Article history:

Received Oct 29, 2021
Revised Mar 25, 2022
Accepted Apr 5, 2022

Keywords:

Acute leukemia detection
Leukemia detection
Lymphoblastic leukemia
Machine learning
Morphological techniques
Support vector machine
classification

ABSTRACT

Acute lymphoblastic leukemia (ALL) has recently been one of the most significant concerns in cancers, especially child and old age. Therefore, crying needs to diagnose leukemia as early as possible, increasing the treatment options and patient survivability. Some basic handicraft leukemia detection processes have been introduced in this arena though these are not so accurate and efficient. The proposed approach has been introduced an automated ALL recognition system from the peripheral blood smear. Initially, the color threshold has been applied to segment lymphocytes blood cells from the blood smear. Some post-processing techniques like morphological operation and watershed have been executed to segment the particular lymphocytes cell. Finally, we used a support vector machine (SVM) classifier to classify the cancerous image frames using a statistical feature vector obtained from the segmented image. The proposed framework has achieved the highest accuracy of 99.21%, the sensitivity of 98.45%, specificity of 99%, the precision of 99%, and F1 score of 99.1%, which has beat existing and common states of art methods. We are confident that the proposed approach will positively impact the ALL detection arena.

Corresponding Author:

Abdullah Al Mamun, Md. Jakir Hossen
Faculty of Engineering and Technology, Multimedia University
Ayer Keroh, Melaka-7540, Malaysia
Email: mamun130203@gmail.com, jakir.hossen@mmu.edu.my

1. INTRODUCTION

The diagnosis of blood cancer is usually considered to be one of the biggest challenges in the healthcare industry as it relies on the hematologist's ability to detect it for a long time. But identifying cancer as soon as possible is very critical for faster responses and improved care options. Computer-aided diagnosis has been introduced to minimize the physician's burden and propeller of data overloading [1]. The primary purpose of computer-aided diagnosis is to detect the abnormality as soon as possible, which is sometimes impossible for the physician manually [2]. Escalante *et al.* [3] have developed an ensemble particle swarm-based model in digitized bone marrow images to recognize acute leukemia particles. This model has achieved better performance for detecting leukemia than other manual procedures. It has reached 97.68% for the binary classification and 94.21% for the multi-categorical classifications [3]. Rawat *et al.* [4] have introduced an intelligent diagnosis system for finding blood cancer which depends on gray level

co-occurrence matrix (GLCM) and graphical features. Support vector machine (SVM) has been examined to classify the normal and abnormal images from the cancer dataset. The model has obtained 89.8% classification accuracy by extracting texture-shape features [4]. In the study, Putzu *et al.* [5] have compared the acute lymphoblastic leukemia (ALL) detection result by examining different classifiers. Their comparison study's highest 93% accuracy with 98% sensitivity has been noted by applying a Gaussian radial kernel-based SVM classifier [5]. Mohapatra *et al.* [6] proposed a cold agglutinin disease (CAD) leukemia diagnostic model based on a fuzzy-segmentation technique and achieved an accuracy of 93% by applying an SVM classifier. Also, normal and blast cells have been classified using an SVM classifier through shape and color features, which has achieved 93.7% accuracy, 92% sensitivity, and 91% specificity in the acute lymphoblastic leukemia-international database-1 (ALL-IDB-1) dataset [7]. Again in study [8], an automated acute lymphoblastic leukemia detection model has been tested on blood smear images to determine the abnormalities.

A decision tree classifier has been applied to get the optimal result to classify and achieve an accuracy of 96.25%, sensitivity of 97.3%, and specificity of 95.35%. Techniques including fluorescence in situ hybridization (FISH), cytochemistry, cytogenetic synthesis, and immunophenotyping have also been introduced to classify leukemia which is very time-consuming [9]. The proposed system [10] can differentiate the leukemia cells containing healthy lymphocytes from the ALL-IDB-1 dataset. Therefore, actual feature extraction of leukemia and typical images is challenging [11], [12]. Shafique *et al.* [13] have proposed color features with an SVM classifier and achieved 93.7% accuracy with low sensitivity and specificity. Tuba *et al.* [14] has introduced shape and texture features and classified leukemia using an SVM classifier. Iterative distance transform has been used to segment the leukemia circles in [15]. In contrast, Al Mamun *et al.* [16] have used fuzzy logic to segment the bleeding portion. Many color threshold segmentation approaches have been applied to differentiate the particular abnormalities from the normal conditions [17]–[20] though these have lower accuracy.

The proposed approach has introduced an efficient automated ALL detection framework from the ALL-IDB dataset. The color threshold approach has been applied to extract the detailed information, and a feature vector has been created by calculating some statistical features. In this article, leukemia detection has been implemented with decent accuracy by using an SVM classifier.

2. RESEARCH METHOD

This proposed model has introduced an automated ALL recognition system from the peripheral blood smear. A complete workflow diagram of the proposed method has been depicted in Figure 1. Initially, the color threshold has been applied to segment lymphocytes blood cells from the blood smear. Some post-processing techniques like morphological operation and watershed have been executed to segment the particular lymphocytes cell. Finally, SVM has been implemented on the feature vector to categorize the cancerous image. A Linux-based environment with version 18.04 bearing having GTX 1080Ti GPU has been used to train and test the proposed model.

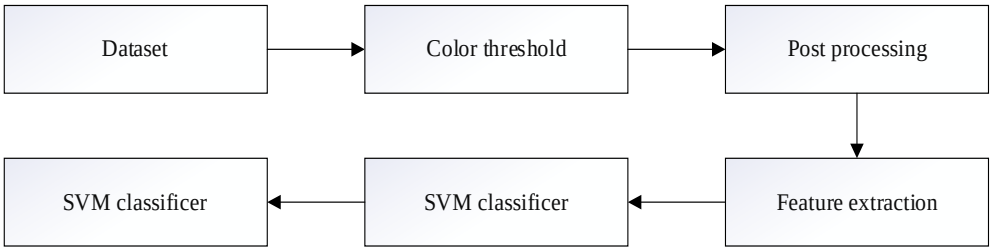
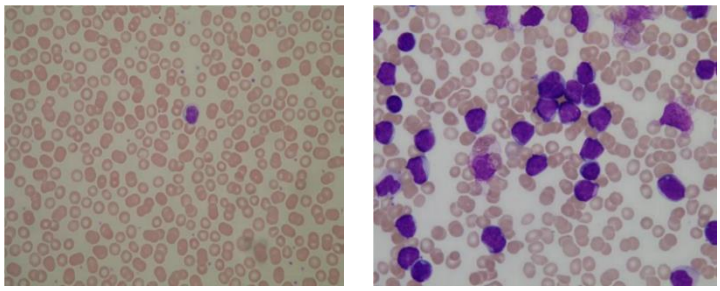


Figure 1. Schematic diagram of the proposed leukemia determination method

2.1. Dataset

Images from ALL-IDB, a publicly available dataset, have been used in this research work [21]. There are 108 image frames in which it has two categorical images like non-cancerous and leukemia. There are 49 image frames of leukemia and 59 image frames of non-cancerous. The resolution of healthy image frames is 2592×1944 and 1712×1368 for leukemia. The sample image frames for healthy and cancerous cells have been shown in Figures 2(a) and 2(b).

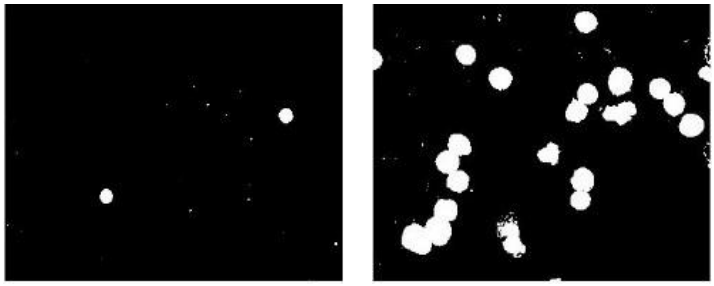


(a) (b)

Figure 2. ALL-IDB sample image frames (a) normal cell and (b) abnormal cell

2.2. Color threshold

The images in the dataset are in the red, green, blue (RGB) color space, and the color intensity for both categories differ from each other. Hence, it is challenging to extract the ALL cells from RGB color space images. The image frames have been converted into YCbCr color space to outperform the intensity of the cancerous and non-cancerous images. A color threshold has been applied to create the mask to extract the informative portion (leukemia) from the image frames. This color mask has segmented the leukemia cells. Also, the non-informative part, excluding leukemia, has been trunked at the same stage, which has been shown in Figures 3(a) and 3(b).



(a) (b)

Figure 3. Segmented image by applying color threshold (a) normal cell and (b) abnormal cell

2.3. Post-processing

The output images from the segmented stage have many noises and unwanted information. These noises may mislead the result of recognition. To suppress the element overflows, a multidimensional image filter has been implemented. Besides, it can compare the value with the target. Watershed transform has been applied, which has reverted with a leveled matrix that differentiates that water region in the image. It detects the watershed ridgelines by assuming the image as the surface where slight pixels mean high superiority and reverse for the dark pixels.

Eventually, the relevant section of the images has been extracted and separated completely from the image background and any other irrelevant objects. This is an efficient technique to spot the object of interest from the image background as well as from other unwanted objects [22]. Having completed the separation of the information from image background and other objects, we have implemented the morphological technique. This technique applies structural elements from strewing, which resizes all the images into the same dimension. During this implementation, each output image pixel is compared with the corresponding pixel of the given image associated with the neighbors [23]. We have formulated a susceptible morphological operation on the given data by appropriately selecting the neighborhood shape and size. In this research work, dilation and erosion have been used with the structural element from strewing. The erosion and dilation were appended and eliminated the pixels of the corresponding object boundaries. The pixel dimension is adjusted depending on shape and size of the structural elements. Finally, after having the morphological operation completed, we got the most suitable informative section from the image. The final segmentation of the images is shown in Figure 4.

Figure 4. Final segmented image of leukemia

2.4. Feature extraction

Feature extraction is a significant step before implementing machine learning-based classification on a dataset. This process is one way of dimension reduction that can efficiently represent a particular section of the image in a vector of features. Our proposed technique has used distinct statistical features, including mean, variance, mode, form factor, eccentricity, entropy, solidity, elongation, compactness, and rectangularity to extract various features and hence to help in machine computation from the segmented section of the image. Since YCbCr color mode has provided the best classification accuracy, feature extraction is recommended in YCbCr color mode for this proposed model.

2.5. Classification

SVM classification is a superior classification model that can also be utilized for classification and regression. Generally, the SVM classifier is a superior method for examining two classes. Recognition of bleeding portion is also a problem of two classes. The SVM classifier provides a training vector which is called a support vector, and it is supposed to be most significant for constructing the decision level. Let a learning data value $pi = p_i(n), n = 1, \dots, N$ is referred to the class level which carries a number of the color feature of M images. The learning data values are a feature vector of the S dimension. To get the feature vector, the function $f(p) = f(k, p)$, k= $[k1\ k2\ \dots\ \dots\dots k_s]^T$ is employed for highly matching the class level y_i. The learning vector pi will convince the provided equation with +1 and -1 for leveling two classes.

$$p_i = \begin{cases} k^T p_i + b \geq +1, & positive\ p_i \\ k^T p_i + b \leq -1, & negative\ p_i \end{cases}$$

The discriminant analysis vector can be defined by the following equation by assuming the seed function $k(x, y)$, and experimental vector c.

$$f(x) = \sum_{i=1}^{M} c_i k(x_i, x) + b$$

SVM classifier can be made more effective in some cases by adapting non-linear seed function [24].

3. RESULTS AND DISCUSSION

Like other classification models, leukemia diagnosis also faced four possible outcomes, which are listed: i) true positive (TP), ii) false positive (FP), iii) true negative aka (TN), and iv) false negative (FN). For this type of classification model, the accuracy metric itself can't justify the method's reliability. So, other relevant parameters, including sensitivity, precision, specificity, F1 score, have been considered to justify the performance of the proposed method. The aforementioned method has recorded the best accuracy of 99.21%, the sensitivity of 98.45%, the specificity value of 99%, the precision value of 99%, and the F1 score of 99.1%, depicted in Figure 5.

Besides, we have tried four different color spaces for developing this model to inspect if color space has any significance on the performance. Figure 6 gives a comparison of the performance in different color spaces, including RGB, hue, saturation, value (HSV), YCbCr, and L*a*b*. From the Figure 6, it is clear that the YCbCr color mode gives the best result in comparison to other color spaces.

Furthermore, the classifier that classifies the images was varied to get the best one from four prospective algorithms. Figure 7 depicts the performance of four classifiers that have been used including quadratic discriminant analysis (QDA), fine Gaussian support vector machine, linear discriminant analysis (LDA), and logistic regression. After analyzing the performance metrics, it is noticeable that SVM is the best classifier for leukemia detection.

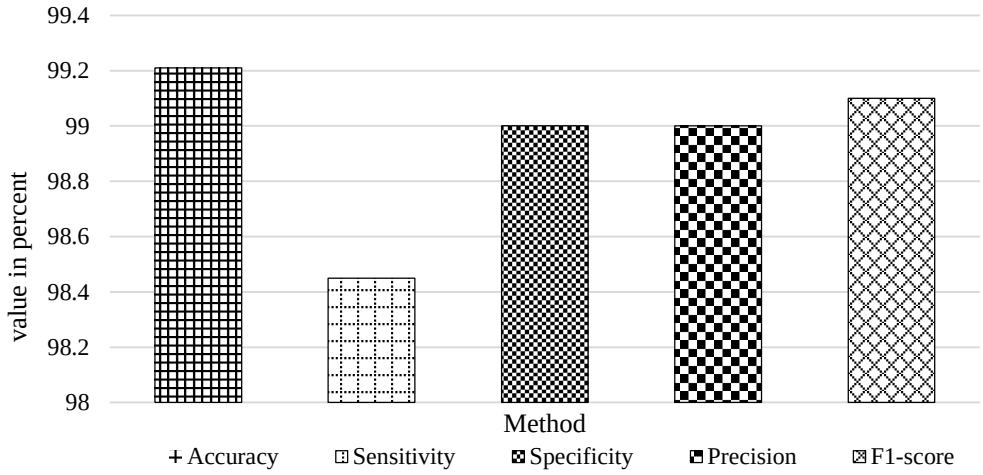


Figure 5. The performance of the proposed leukemia detection model

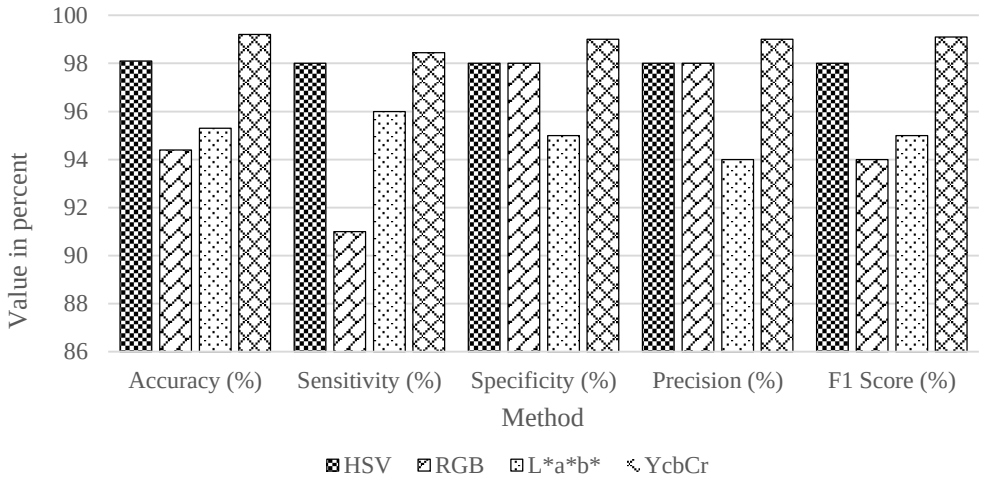


Figure 6. The comparison results in different color space

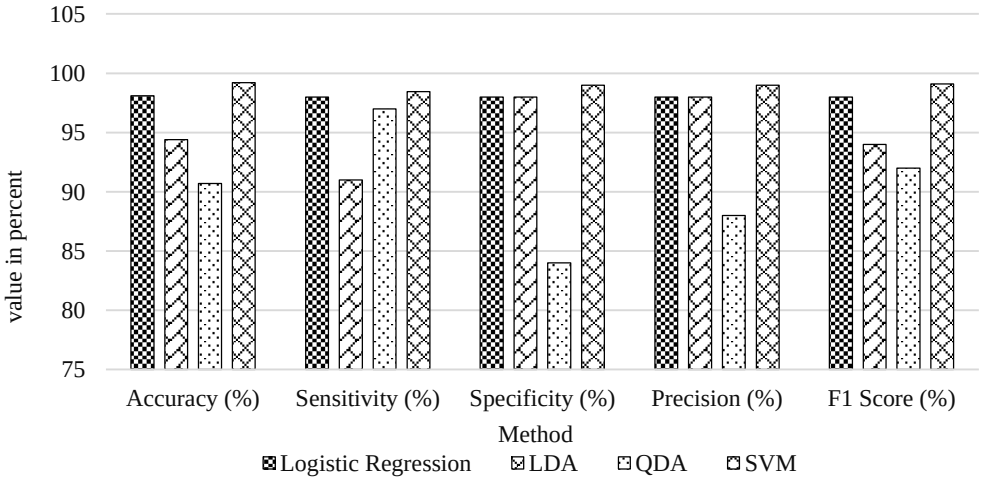


Figure 7. The comparison results in different classifiers

Even though much research is going on leukemia determination, our proposed method is unique and supposed to contribute to this research domain. A comparison between the existing method and our proposed method is shown in Table 1. From this table, we can notice that the proposed method beats the existing method in terms of accuracy, sensitivity, and specificity. Besides, this research work has also analyzed a few other relevant performance metrics, including precision, negative-predicted value, and F1 score. In a word, this proposed technique is a more accurate and reliable model for leukemia detection.

Table 1. A comparison study on different existing methods

Methods	Classifier	Acc (%)	Sen (%)	Spe (%)
Putzu *et al*. [5]	SVM	93	98	-
Li *et al*. [7]	SVM	93.3	92	91
El Houby [8]	Decision tree	96.25	97.3	95.35
Umamaheswari and Geetha [25]	KNN	96.25	95	97
Proposed method	**SVM**	**99.21**	**98.45**	**99**

4. CONCLUSION

ALL is one of the fatal diseases at present. An accurate and reliable diagnostic method of ALL is the prerequisite of successful ALL is a treatment. The major concern of the proposed method is to diagnose ALL more accurately and efficiently. Consequently, the proposed approach has introduced an efficient automated ALL detection framework from the ALL-IDB dataset. The color threshold approach has been applied to extract the particular information, and a feature vector has been created by calculating some statistical features. Here, SVM based machine learning classifier has been employed to classify leukemia from the normalities. The proposed framework has achieved the highest accuracy of 99.21%, sensitivity of 98.45%, specificity of 99%, precision of 99%, and F1 score of 99.1%, which outperforms some existing research. The result can be made robust by examining different advanced classifiers and features.

ACKNOWLEDGEMENTS

We would like to show our appreciation and thanks to Multimedia University, Melaka, Malaysia for supporting this research work.

REFERENCES

[1] C.-M. Chen, Y.-H. Chou, N. Tagawa, and Y. Do, "Computer-aided detection and diagnosis in medical imaging," *Computational and Mathematical Methods in Medicine*, vol. 2013, pp. 1–2, 2013, doi: 10.1155/2013/790608.
[2] B. Halalli and A. Makandar, "Computer aided diagnosis-medical image analysis techniques," in *Breast Imaging*, InTech, 2018.
[3] H. J. Escalante *et al.*, "Acute leukemia classification by ensemble particle swarm model selection," *Artificial Intelligence in Medicine*, vol. 55, no. 3, pp. 163–175, Jul. 2012, doi: 10.1016/j.artmed.2012.03.005.
[4] J. Rawat, A. Singh, H. S. Bhadauria, and J. Virmani, "Computer aided diagnostic system for detection of leukemia using microscopic images," *Procedia Computer Science*, vol. 70, pp. 748–756, 2015, doi: 10.1016/j.procs.2015.10.113.
[5] L. Putzu, G. Caocci, and C. Di Ruberto, "Leucocyte classification for leukaemia detection using image processing techniques," *Artificial Intelligence in Medicine*, vol. 62, no. 3, pp. 179–191, Nov. 2014, doi: 10.1016/j.artmed.2014.09.002.
[6] S. Mohapatra, S. S. Samanta, D. Patra, and S. Satpathi, "Fuzzy based blood image segmentation for automated leukemia detection," in *2011 International Conference on Devices and Communications (ICDeCom)*, Feb. 2011, pp. 1–5, doi: 10.1109/ICDECOM.2011.5738491.
[7] Y. Li, R. Zhu, L. Mi, Y. Cao, and D. Yao, "Segmentation of white blood cell from acute lymphoblastic leukemia images using dual-threshold method," *Computational and Mathematical Methods in Medicine*, pp. 1–12, 2016, doi: 10.1155/2016/9514707.
[8] E. M. F. El Houby, "Framework of computer aided diagnosis systems for cancer classification based on medical images," *Journal of Medical Systems*, vol. 42, no. 8, Aug. 2018, doi: 10.1007/s10916-018-1010-x.
[9] S. Mohapatra, D. Patra, and S. Satpathi, "Image analysis of blood microscopic images for acute leukemia detection," in *2010 International Conference on Industrial Electronics, Control and Robotics*, Dec. 2010, pp. 215–219, doi: 10.1109/IECR.2010.5720171.
[10] A. Genovese, M. S. Hosseini, V. Piuri, K. N. Plataniotis, and F. Scotti, "Acute lymphoblastic leukemia detection based on adaptive unsharpening and deep learning," in *ICASSP 2021-2021 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*, Jun. 2021, pp. 1205–1209, doi: 10.1109/ICASSP39728.2021.9414362.
[11] W. Xie, Y. Li, and Y. Ma, "Breast mass classification in digital mammography based on extreme learning machine," *Neurocomputing*, vol. 173, pp. 930–941, Jan. 2016, doi: 10.1016/j.neucom.2015.08.048.
[12] A. K. Mohanty, M. R. Senapati, and S. K. Lenka, "Retracted article: An improved data mining technique for classification and detection of breast cancer from mammograms," *Neural Computing and Applications*, vol. 22, no. S1, pp. 303–310, May 2013, doi: 10.1007/s00521-012-0834-4.
[13] S. Shafique, S. Tehsin, S. Anas, and F. Masud, "Computer-assisted acute lymphoblastic leukemia detection and diagnosis," in *2019 2nd International Conference on Communication, Computing and Digital systems (C-CODE)*, Mar. 2019, pp. 184–189, doi: 10.1109/C-CODE.2019.8680972.
[14] E. Tuba, I. Strumberger, N. Bacanin, D. Zivkovic, and M. Tuba, "Acute lymphoblastic leukemia cell detection in microscopic digital images based on shape and texture features," in *Lecture Notes in Computer Science (including subseries Lecture Notes in*

Artificial Intelligence and Lecture Notes in Bioinformatics), vol. 11656 LNCS, Springer Verlag, 2019, pp. 142–151.
[15] N. S. Fatonah, H. Tjandrasa, and C. Fatichah, "Automatic leukemia cell counting using iterative distance transform for convex sets," *International Journal of Electrical and Computer Engineering (IJECE)*, vol. 8, no. 3, pp. 1731–1740, Jun. 2018, doi: 10.11591/ijece.v8i3.pp1731-1740.
[16] A. Al Mamun, P. P. Em, T. Ghosh, M. M. Hossain, M. G. Hasan, and M. G. Sadeque, "Bleeding recognition technique in wireless capsule endoscopy images using fuzzy logic and principal component analysis," *International Journal of Electrical and Computer Engineering (IJECE)*, vol. 11, no. 3, pp. 2688–2695, Jun. 2021, doi: 10.11591/ijece.v11i3.pp2688-2695.
[17] A. Al Mamun, M. S. Hossain, P. P. Em, A. Tahabilder, R. Sultana, and M. A. Islam, "Small intestine bleeding detection using color threshold and morphological operation in WCE images," *International Journal of Electrical and Computer Engineering (IJECE)*, vol. 11, no. 4, pp. 3040–3048, Aug. 2021, doi: 10.11591/ijece.v11i4.pp3040-3048.
[18] A. Al Mamun, M. S. Hossain, M. M. Hossain, and M. G. Hasan, "Discretion way for bleeding detection in wireless capsule endoscopy images," in *2019 1st International Conference on Advances in Science, Engineering and Robotics Technology (ICASERT)*, May 2019, pp. 1–6, doi: 10.1109/ICASERT.2019.8934589.
[19] A. Al Mamun and M. S. Hossain, "Ulcer detection in image converted from video footage of wireless capsule endoscopy," in *2019 1st International Conference on Advances in Science, Engineering and Robotics Technology (ICASERT)*, May 2019, pp. 1–4, doi: 10.1109/ICASERT.2019.8934597.
[20] M. S. Hossain, A. Al Mamun, T. Ghosh, M. G. Hasan, M. M. Hossain, and A. Tahabilder, "Ulcer detection in wireless capsule endoscopy using locally computed features," in *Lecture Notes of the Institute for Computer Sciences, Social-Informatics and Telecommunications Engineering, LNICST*, vol. 325 LNICST, no. Cd, 2020, pp. 491–502.
[21] R. D. Labati, V. Piuri, and F. Scotti, "All-IDB: The acute lymphoblastic leukemia image database for image processing," in *2011 18th IEEE International Conference on Image Processing*, Sep. 2011, pp. 2045–2048, doi: 10.1109/ICIP.2011.6115881.
[22] F. Meyer, "Topographic distance and watershed lines," *Signal Processing*, vol. 38, no. 1, pp. 113–125, Jul. 1994, doi: 10.1016/0165-1684(94)90060-4.
[23] R. C. Gonzalez, R. E. Woods, and S. L. Eddins, *Digital image processing using MATLAB*. USA: Prentice-Hall, Inc., 2003.
[24] T. Ghosh, S. A. Fattah, and K. A. Wahid, "Automatic bleeding detection in wireless capsule endoscopy based on RGB pixel intensity ratio," in *2014 International Conference on Electrical Engineering and Information and Communication Technology*, Apr. 2014, pp. 1–4, doi: 10.1109/ICEEICT.2014.6919173.
[25] D. Umamaheswari and S. Geetha, "A framework for efficient recognition and classification of acute lymphoblastic leukemia with a novel customized-KNN classifier," *Journal of Computing and Information Technology*, vol. 26, no. 2, pp. 131–140, Jun. 2018, doi: 10.20532/cit.2018.1004123.

BIOGRAPHIES OF AUTHORS

Abdullah Al Mamun has received a B.Sc. degree in Electrical and Electronic Engineering from Pabna University of Science and Technology in 2018. Now he is pursuing M.Eng.Sc. at Multimedia University (MMU) in the Faculty of Engineering and Technology since 2019. He is also working as a lecturer (on study leave) since November 2019 in EEE at Feni University, Feni, Bangladesh. His research interest includes computer vision, image processing, signal processing, deep learning and machine learning. He can be reached at the email: mamun130203@gmail.com.

Md. Jakir Hossen is graduated in Mechanical Engineering from the Dhaka University of Engineering and Technology (Former Name is BIT Dhaka,1997), Higher Diploma in Computer Science and Engineering (1999), Masters in Communication and Network Engineering from Universiti Putra Malaysia (2003), and Ph.D. in Smart Technology and Robotic Engineering at Universiti Putra Malaysia (2012). He is currently a Senior Lecturer at the Faculty of Engineering and Technology, Multimedia University, Melaka, Malaysia. His research interests are in the area of artificial intelligence, data analytics, robotics and control. He can be contacted at email: jakir.hossen@mmu.edu.my.

Anik Tahabilder is currently pursuing his Ph.D. in Computer Science at Wayne State University, Detroit, MI-48202, USA. He had his master's degree in Engineering Technology from Western Carolina University, Cullowhee, NC-28723, United State. He also got a bachelor's degree in Electrical and Electronic Engineering from Pabna University of Science and Technology, Pabna-6600, Bangladesh. His research interest includes machine learning, deep learning, big data and blockchain. He can be contacted at email: tahabilderanik@gmail.com.

Ahmmad Musha is pursuing a B.Sc. degree in Electrical and Electronic Engineering from Pabna University of Science and Technology (PUST), Pabna, Bangladesh, since 2016. His research interest includes deep learning, biomedical engineering, robotics and automation. He can be contacted at email: mushaxyz@gmail.com.

Rehnuma Hasnat is pursuing a B.Sc. degree in Electrical and Electronic Engineering from Pabna University of Science and Technology, Pabna, Bangladesh. Her research interest includes biomedical engineering, deep learning, and machine learning, image processing, signal processing, and smart grid. She can be contacted at email: rehnumahasnat@gmail.com.

Sohag Kumar Saha has received his B.Sc. degree in Electrical and Electronic Engineering from Pabna University of Science and Technology (PUST) in 2014. Currently, he is pursuing PhD in Electrical and Computer Engineering at Tennessee Technological University, Cookeville, TN, USA. He has also worked more than 5 years as a Sub-Divisional Engineer (SDE) at Ashuganj Power Station Company Ltd., Brahmanbaria, Bangladesh. His research interests include power systems, power electronics, smart grid systems, machine learning, deep learning and artificial intelligence. He can be contacted at email: sohag@ieee.org.