

Leukaemia Identification based on Texture Analysis of Microscopic Peripheral Blood Images using Feed-Forward Neural Network

Wahyu Tri Puspitasari¹, Dina Zatusiva Haq², Dian Candra Rini Novitasari^{3*}

^{1,2,3}*Department of Mathematics, Faculty of Science and Teknologi, UIN Sunan Ampel*

**diancrini@uinsby.ac.id*

ABSTRACT

Leukaemia is very dangerous because it includes liquid tumour that it cannot be seen physically and is difficult to detect. Alternative detection of Leukaemia using microscopy can be processed using a computing system. Leukemia disease can be detected by microscopic examination. Microscopic test results can be processed using machine learning for classification systems. The classification system can be obtained using Feed-Forward Neural Network. Extreme Learning Machine (ELM) is a neural network that has a feedforward structure with a single hidden layer. ELM chooses the input weight and hidden neuron bias at random to minimize training time based on the Moore Penrose Pseudoinverse theory. The classification of Leukaemia is based on microscopic peripheral blood images using ELM. The classification stages consist of pre-processing, feature extraction using GLRLM, and classification using ELM. This system is used to classify Leukaemia into three classes, that is acute lymphoblastic Leukaemia, chronic lymphoblastic Leukaemia, and not Leukaemia. The best results were obtained in ten hidden nodes with an accuracy of 100%, a precision of 100%, a withdrawal of 100%.

Keywords: Leukaemia, Feature Extraction, Classification, Microscopic Peripheral Blood, Texture Analysis, Feed-Forward Neural Network.

1. INTRODUCTION

Leukaemia and Lymphoma Society states that Leukaemia is a deadly disease for all ages [1]. Leukemia is a blood disorder in the bone marrow that attacks blood-forming cells [2]. Leukemia is in the form of a liquid tumor so that it cannot be seen physically and is difficult to detect [3]. If the bone marrow produces excessive or abnormal white blood cells, it can affect the white blood cells to protect the body, which is usually called leukaemia. Leukaemia, based on its length of time, is classified into acute leukaemia and chronic leukaemia [4].

World Health Organization (WHO) states that there are approximately 150,000 people with leukaemia of acute lymphocytic type in each year with a relatively high mortality rate in Indonesia [5]. However, early detection is difficult because the general symptoms of leukaemia are not quickly recognised [6]. Alternative detection of Leukaemia using microscopy can be processed using a computing system. Leukemia disease can be detected by microscopic examination [7]. Microscopic examination counts the number of blood cells in the body. The computing process in the leukemia diagnosis system is carried out by a classification process based on the results of microscopic examinations [8].

Wahyu Tri Puspitasari, Dina Zatusiva Haq, Dian Candra Rini Novitasari
Leukaemia Identification based on Texture Analysis of Microscopic Peripheral Blood
Images using Feed-Forward Neural Network

Previously, several studies have been conducted regarding the classification of Leukaemia. Rangole used neural networks for Leukaemia detection and obtained the best results with an accuracy of 83.33% [9]. Jyoti Rawat also implemented another method to detect leukaemia, that is feature extraction using the Gray-Level Co-Occurrence Matrix (GLCM) which obtained the best results with an accuracy of 89.8% [10]. Sendren proved that the performance of the Gray-Level Run Length Matrix (GLRLM) was better than the GLCM method [9]. The difference between GLCM and GLRLM was found in the matrix and statistical characteristics obtained. These statistical characteristics would be used as input in classifications using machine learning [12], [13].

In recent years, there have been developments of machine learning, one of which is an Feed-Forward Neural Network. Extreme Learning Machine (ELM) is a neural network that has a feedforward structure with a single hidden layer [14]. ELM is widely used to complete classification. Nemissi uses ELM for the classification of breast cancer with excellent results. Besides, ELM has also been used to solve problems such as forecasting the inflation rate in Indonesia [14], the detection of dengue fever [15], the diagnosis of breast cancer [16], and others showing results with an average accuracy rate above 90%. Gowthaman compares SVM with ELM. From several experiments, it is concluded that ELM is better than SVM in terms of results and learning speed [15].

2. MATERIAL AND METHODS

The classification system used microscopic peripheral blood image data to identify leukemia into three classes: acute, chronic, and non-leukemia, using the ELM method. Microscopic peripheral blood images were obtained from the Acute Lymphoblastic Leukaemia Image Database for Image Processing (ALL-IDB). This dataset consisted of 161 microscopic peripheral blood images with 49 Acute Lymphoblastic Leukaemia blood image data, 53 chronic lymphoblastic Leukaemia image data, and 59 non-leukaemia blood image data [18]. There are three stages in this system: pre-processing, feature extraction, and classification.

The Pre-processing stage is used to improve image quality before the feature is taken. Microscopic peripheral blood data is an RGB image. Next, RGB images are converted into grayscale images before the feature extraction process. The next step is histogram equalization to correct the intensity because the image has an uneven intensity. Then, the filtering process uses a median filter. The purpose of filtering is to eliminate noise contained in the image. After improving image quality, the next process is feature extraction using GLRLM to obtain the features contained in the image.

The first step in feature extraction using GLRLM is to construct a run-length matrix using $p(r, t | \theta)$ where r is the grey level, t is the number of occurrences of run length, and θ is the direction of direction orientation. These features are used as input for the classification process.

The next stage is the division of data into two, that is training data and testing data. Data sharing uses the k-fold cross-validation. Training data is intended for the learning process of classification systems. Testing data is intended for testing the results of the classification. This learning process uses ELM. The input matrix in ELM is processed with weights and biases that have been initialized in advance to obtain the output of the hidden layer. The activation function is used to obtain the output of the hidden

layer. The hidden layer output is an H matrix obtained from Equation 10. Then, the pseudo-inverse model is calculated. The pseudo-inverse model is used to calculate the value of beta weights (β) using equation 13. After obtaining β , the next process is training. The training is conducted using testing data with the same weight and bias values as the learning process. After that, the H matrix from the testing data is calculated. The H matrix is multiplied by β to obtain the classification results. The illustration of research step is shown in floxchart Figure 1.

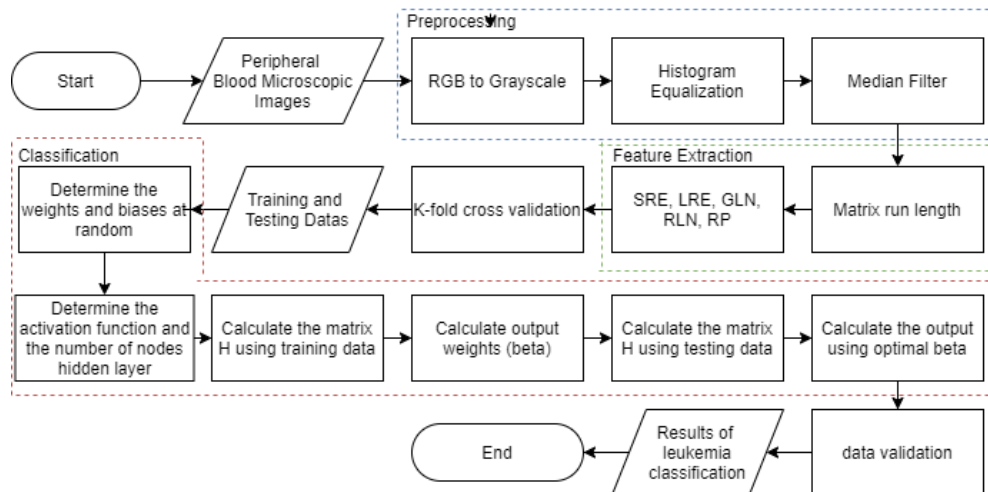


FIGURE 1. Research Flowchart

2.1 GRAY LEVEL RUN-LENGTH MATRIX

Gray-Level Run Length Matrix is one of the methods used to extract features contained in an image to obtain statistical functions or features included in the model by estimating pixels with the same degree of grey [19]. If it is known that the run-length matrix with the matrix element $p(r, t | \theta)$ gives the total number of occurrences of run that has length t of gray-level r in a given direction orientation θ [20]. The direction orientation is formed with four shift directions and Galloway said that several types of statistical features could be extracted from the run-length matrix, that is[21]:

2.1.1 SHORT RUNS EMPHASIS (SRE)

Short Runs Emphasis (SRE) is a feature of measuring the distribution of short runs in the image. SRE is small if the image has a smooth texture, while large SRE means that the image has a rough surface [22].

$$SRE = \sum_{r=1}^K \sum_{t=1}^L \frac{p(r,t|\theta)}{t^2} / \sum_{r=1}^K \sum_{t=1}^L p(r,t|\theta) \quad (1)$$

Description:

$p(r, t | \theta)$ The run-length matrix
 r Grey level of image

T	Length of image
a_i	Peripheral blood microscopic images
b_j	The bias weight of hidden neuron

2.1.2 LONG RUNS EMPHASIS (LRE)

Long Runs Emphasis (LRE) is a feature of measuring the long-run distribution in the image. LRE is large if the texture is smooth but small LRE means rough surface [9s].

$$LRE = \sum_{r=1}^K \sum_{t=1}^L t^2 p(r, t|\theta) / \sum_{r=1}^K \sum_{t=1}^L p(r, t|\theta) \quad (2)$$

2.1.3 GRAY LEVEL NON-UNIFORMITY (GLN)

Gray Level Non-uniformity (GLN) is a feature that shows the degree of gray in all parts of the image. The smaller the GLN value, the more evenly the level of gray in an image.

$$GLN = \sum_{r=1}^K (\sum_{t=1}^L p(r, t|\theta))^2 / \sum_{r=1}^K \sum_{t=1}^L p(r, t|\theta) \quad (3)$$

2.1.4 RUN LENGTH NON-UNIFORMITY (RLN)

Run Length Non-uniformity (RLN) measures the run-length equation throughout the image. RLN is small if the run length is the same as the whole image.

$$RLN = \sum_{t=1}^L (\sum_{r=1}^K p(r, t|\theta))^2 / \sum_{r=1}^K \sum_{t=1}^L p(r, t|\theta) \quad (4)$$

2.1.5 RUN PERCENTAGE (RP)

Run Percentage (RP) measures the uniformity and distribution of the run of an image in a particular direction. RP is large if the run length is 1 for all gray levels for a specific direction.

$$RP = \frac{1}{N} \sum_{r=1}^K \sum_{t=1}^L p(r, t|\theta) \quad (5)$$

2.2 EXTREME LEARNING MACHINE (ELM)

Extreme Learning Machine (ELM) is part of a feedforward artificial neural network, which is also known as Single Hidden Layer Forward Feed Neural Networks [24]. In conventional artificial neural networks, updating the input weights and hidden neurons requires a complex computational process with a long enough time for iterations to optimize the weights [25]. ELM chooses the input weight and hidden neuron bias at random to minimize training time based on the Moore Penrose Pseudoinverse theory. In this way, ELM can work in complex datasets and with minimal time [26].

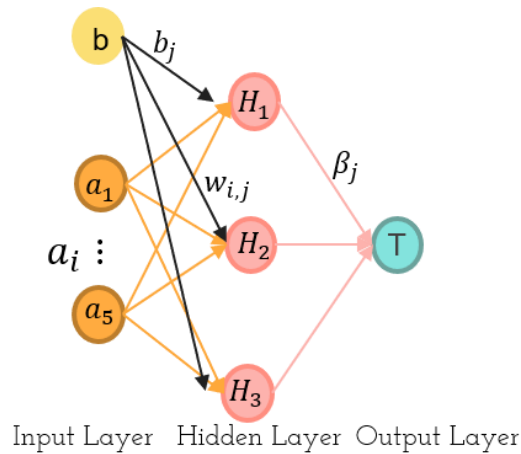


FIGURE 2. ELM Architecture

2.2.1 INITIALIZE

Initialize a , $f(u)$, w and b . There are N samples (a_i, y_i) with $a_i = [a_{i1}, a_{i2}, \dots, a_{in}]' \in \mathbb{R}_n$ and $y_i = [y_{i1}, y_{i2}, \dots, y_{in}]' \in \mathbb{R}_n$. n = number of parameters and with m hidden neurons and the activation function $f(u)$. Where $w_{.j} = [w_{11}, w_{22}, \dots, w_{nm}]'$ is a weight vector that connects nodes in the hidden layer to nodes from the input layer and b is a bias weight [27].

2.2.2 CALCULATE HIDDEN LAYER OUTPUT (H)

$$H = \begin{bmatrix} f(a_{1,:}, w_{1,1} + b_1) & \cdots & f(a_{1,:}, w_{1,m} + b_m) \\ \vdots & \ddots & \vdots \\ f(a_{N,:}, w_{N,1} + b_1) & \cdots & f(a_{N,:}, w_{N,m} + b_m) \end{bmatrix} \quad (6)$$

2.2.3 CALCULATE WEIGHT OUTPUT

$$\beta = H^\dagger T \quad (7)$$

Description:

$f(u)$	Activation function
w_j	Weight vector that connects nodes on the hidden layer to w with nodes from the input layer
y_i	Label of images
N	Number of data inputs
n	Number of parameters
m	Number of hidden layers

Wahyu Tri Puspitasari, Dina Zatusiva Haq, Dian Candra Rini Novitasari
Leukaemia Identification based on Texture Analysis of Microscopic Peripheral Blood
Images using Feed-Forward Neural Network

H	Hidden Layer Output
H^+	Pseudoinverse of matrix A
T	Output target

3. RESULT AND DISCUSSION

In this study, the Leukaemia classification system was designed using Extreme Learning Machine based on microscopic peripheral blood images. The following data samples from various levels of lymphoblastic Leukaemia are shown in Figure 3. There are three stages in this system: pre-processing, feature extraction, and classification.

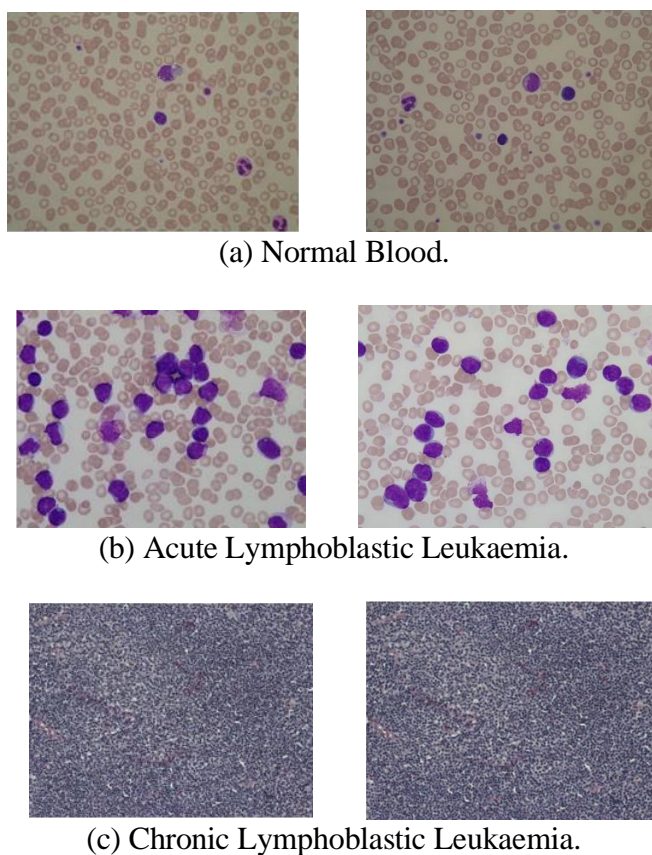


FIGURE 3. Samples of Microscopic Peripheral Blood Images

The microscopic peripheral blood image is used as an input in the Leukaemia classification system using the Extreme Learning Machine. The stages of this research consist of pre-processing, feature extraction, and classification. The first stage is pre-processing. The microscopic peripheral blood image is changed to grayscale. The grayscale process aims to take grayscale images so that it can facilitate the pre-processing. In the preprocessing stage, it is applied histogram equalization process to enhancing the light intensity of the image.

The histogram equalization process is then performed. The purpose of histogram equalization is to level the light intensity of the image. The reason for the existence of a histogram equalization process is because the data processed has varying light intensities. The next process is median filter. A median filter is used to remove noise

contained in the image. The improved quality images are used as input in the feature extraction process. The following is the after pre-processing image samples shown in Figure 4.

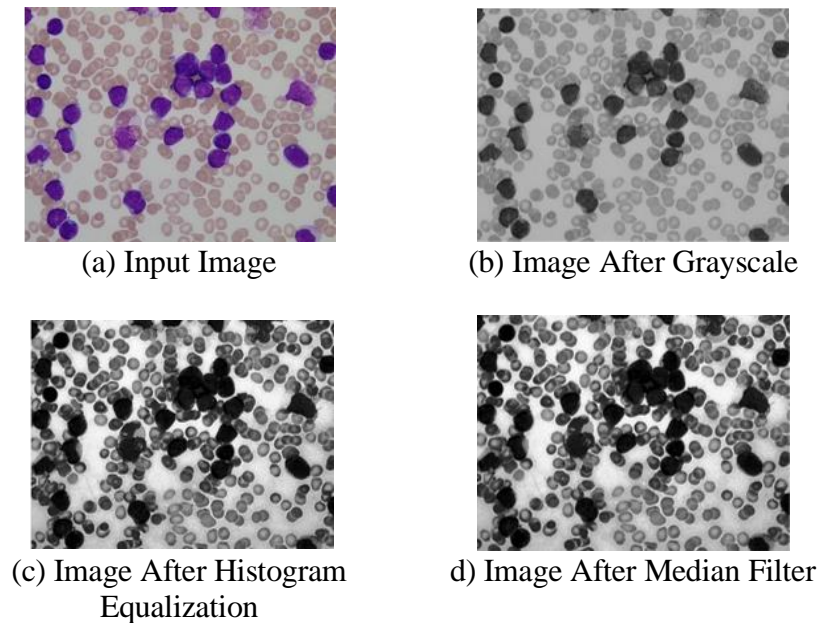


FIGURE 4. Pre-processing Result

The first step of feature extraction using GLRLM is to create a run-length matrix. From the matrix, several features will be obtained, including SRE, LRE, GLN, RLN and RP. In this study, a run-length matrix was formed from each direction orientation of 0° . The following feature samples from six images of the 0° orientation are presented in Table 1.

Based on Table 1, it is shown that the LRE value of non-leukaemia blood image is greater than that of the ALL patient's blood image and the ALL patient's blood image has a higher LRE value than the CLL patient's blood image. The most significant LRE value obtained is the non-leukaemia blood image which means that this image has a rough texture. This corresponds to a small SRE. While the LRE value of CLL patient images has the least value, which is around 2.06. It means that the CLL patient's blood images have smooth textures. This is because the number of the long run is small.

TABLE 1.
Result of Feature Extraction Samples Using GLRLM

Image to-	SRE	LRE	GLN	RLN	RP	Class
1	0.550	9.417	658321.805	67025.136	3.335	Normal
2	0.544	9.991	638747.423	69185.604	3.291	Normal
3	0.634	7.653	437172.782	27380.224	2.658	ALL
4	0.625	7.936	417708.780	28055.258	2.609	ALL
5	0.865	2.068	813890.113	19528.020	3.256	CLL
6	0.857	2.100	774658.421	19817.787	3.222	CLL

Wahyu Tri Puspitasari, Dina Zatusiva Haq, Dian Candra Rini Novitasari
Leukaemia Identification based on Texture Analysis of Microscopic Peripheral Blood
Images using Feed-Forward Neural Network

The next step is classification. ELM is used for the classification process of the data obtained from the previous image processing. The classification process consists of two stages, that is training stage and testing stage. The training stage is to build the best model while the testing stage is to test the accuracy of the model formed. This study used 128 data for the training stage and 33 data for the testing stage. The training and testing conducted used the ELM classification method. In this study, Leukaemia classification used five parameters obtained from the feature extraction using GLRLM, that is SRE, LRE, RLN, GLN, and RP grouped into three classes, that is non-leukaemia (normal), Acute Lymphoblastic Leukaemia (ALL) and Chronic Lymphoblastic Leukaemia (CLL). The value of the features obtained has a different range so that the normalization process is required to facilitate the training stage and the testing stage. The training stage is done by finding the H matrix using random weights and biases. Next, the output weight (β) is calculated. The output weight of the training process is called optimal beta. Then, the network output based on testing data using the optimal beta is calculated. The output is validated using several criteria, that is accuracy, precision, and recall. The following is the results of Leukaemia classification using ELM shown in Table 2 and visualize in Figure 5.

TABLE 2.
Experiment Result

Angle of GLRLM	Hidden Nodes	Accuracy	Precision	Recall	Average Accuracy
0	10	100.00	100.00	100.00	97.93
	11	96.97	96.97	97.22	
	12	96.88	96.97	97.22	
	13	100.00	100.00	100.00	
	14	96.88	96.97	96.97	
	15	96.88	96.97	96.67	
45	10	96.97	97.44	96.67	98.36
	11	96.97	96.97	97.22	
	12	100.00	100.00	100.00	
	13	100.00	100.00	100.00	
	14	96.97	97.44	96.67	
	15	99.22	99.17	99.29	
90	10	96.88	97.44	96.67	97.41
	11	93.75	93.89	93.64	
	12	96.97	97.44	96.67	
	13	100.00	100.00	100.00	
	14	100.00	100.00	100.00	
	15	96.88	96.97	97.22	
135	10	100.00	100.00	100.00	97.44
	11	93.94	93.89	93.89	
	12	96.97	96.97	97.22	
	13	100.00	100.00	100.00	
	14	96.97	96.97	96.97	
	15	96.77	97.44	96.30	

Table 2 shows the evaluation results of corner trials on GLRLM and hidden nodes on EL. At the GLRLM angle of 0 degrees, the highest accuracy reaches 100% accuracy with the number of hidden nodes 10 and 13. The number of hidden nodes 13 reaches the highest accuracy of 100% at each angle. The GLRLM implementation with angles of 45, 90, and 135 degrees and the number of hidden nodes 12, 14, and 10 also has a perfect system performance with accuracy reaching 100%. The average system evaluation results at each angle are visualized in Figure 5.

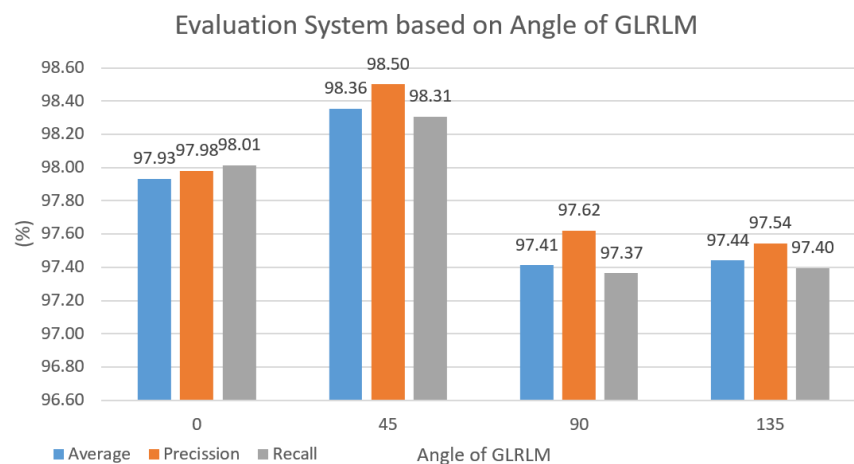


FIGURE 5. Classification Result Using ELM

Based on Figure 4, the average accuracy results at each angle indicate that the feature extraction results using the GLRLM method can represent microscopic peripheral blood images well. The best average accuracy, precision, and recall results from all trials were obtained using the 45-degree angle GLRLM method. The average value of each accuracy, precision, and recall at an angle of 45 degrees is 98.36%, 98.50%, and 98.31%. The second order of classification systems is based on the value of accuracy, namely by using an angle of 0 degrees which is 0.43% lower than the average accuracy at an angle of 45. While the lowest result is a classification system using image features produced by the GLRLM method at an angle of 90 degrees with the average value of each -accuracy, precision, and recall are 97.41%, 97.62%, and 97.37%.

4. CONCLUSION

The Leukaemia classification model using extreme learning machine based on GLRLM features with 161 microscopic peripheral blood data showed good performance. The results of Leukaemia classification into three classes, that is non-leukaemia, ALL and CLL obtained the best results using 0° orientation with the number of neurons in the hidden layer as much as 10 with an accuracy of 100%, a precision of 100% and a recall of 100%. This model had a fast calculation time of 0.2 seconds. This showed that the extreme learning machine could classify the data with an excellent performance and fast time. Additional data was required to generate a better model.

REFERENCES

- [1] K. J. Norsworthy *et al.*, “FDA approval summary: glasdegib for newly diagnosed acute myeloid leukemia,” *Clin. Cancer Res.*, vol. 25, no. 20, pp. 6021–6025, 2019.
- [2] T. Dharani and S. Hariprasath, “Diagnosis of leukemia and its types using digital image processing techniques,” in *2018 3rd International Conference on Communication and Electronics Systems (ICCES)*, 2018, pp. 275–279.
- [3] S. Shafique and S. Tehsin, “Acute Lymphoblastic Leukemia Detection and Classification of Its Subtypes Using Pretrained Deep Convolutional Neural Networks,” vol. 17, pp. 1–7, 2018, doi: 10.1177/1533033818802789.
- [4] P. Pandey, “Cancer of White Blood Cells: Blood Cancer,” *J Tum Res Rep*, vol. 6, p. 141, 2021.
- [5] WHO, “World Health Organization,” 2018. .
- [6] M. Ghaderzadeh, F. Asadi, A. Hosseini, D. Bashash, H. Abolghasemi, and A. Roshanpour, “Machine learning in detection and classification of leukemia using smear blood images: a systematic review,” *Sci. Program.*, vol. 2021, 2021.
- [7] M. E. Billah and F. Javed, “Bayesian Convolutional Neural Network-based Models for Diagnosis of Blood Cancer,” *Appl. Artif. Intell.*, vol. 36, no. 1, p. 2011688, 2022.
- [8] R. Sigit, M. M. Bachtiar, and M. I. Fikri, “Identification of Leukemia diseases based on microscopic human blood cells using image processing,” in *2018 International Conference on Applied Engineering (ICAE)*, 2018, pp. 1–5.
- [9] J. Rangole, “Detection of Leukemia in Microscopic Images Using Image Processing,” in *International Conference on Communication and Signal Processing*, 2019, no. April 2014, doi: 10.1109/ICCSP.2014.6949840.
- [10] J. Rawat, A. Singh, H. S. Bhaduria, and J. Virmani, “Computer Aided Diagnostic System for Detection of Leukemia,” *Int. Conf. Eco-friendly Comput. Commun. Syst.*, vol. 70, pp. 748–756, 2015, doi: 10.1016/j.procs.2015.10.113.
- [11] A. Harshavardhan, S. Babu, and T. Venugopal, “Analysis of feature extraction methods for the classification of brain tumor detection,” *Int. J. Pure Appl. Math.*, vol. 117, no. 7, pp. 147–155, 2017.
- [12] D. C. R. e Novitasari and T. Al., “Application of Feature Extraction for Breast Cancer using One Order Statistic, GLCM, GLRLM, and GLDM,” *Adv. Sci. Technol. Eng. Syst. J.*, vol. 4, no. 4, pp. 114–120, 2019.
- [13] A. Z. Foeady, D. C. R. Novitasari, A. H. Asyhar, and M. Firmansjah, “Automated Diagnosis System of Diabetic Retinopathy Using GLCM Method and SVM Classifier,” *Proceeding Electr. Eng. Comput. Sci. Informatics*, vol. 5, no. 1, pp. 154–160, 2018.
- [14] A. N. Alfiyatin, W. F. Mahmudy, C. F. Ananda, and Y. P. Anggodo, “Penerapan Extreme Learning Machine (ELM) untuk Peramalan Laju Inflasi di Indonesia,” *J. Teknol. Inf. dan Ilmu Komput.*, vol. 6, no. 2, p. 179, 2019, doi: 10.25126/jtiik.201962900.
- [15] S. Suhaeri, N. M. Nawi, and M. Fathurahman, “Early Detection of Dengue Disease Using Extreme Learning Machine,” *Int. J. Adv. Sci. Eng. Inf. Technol.*, vol. 8, no. 5, pp. 2219–2224, 2018.
- [16] M. Nemissi, H. Salah, and H. Seridi, “Breast cancer diagnosis using an

- enhanced Extreme Learning Machine based-Neural Network,” in *International Conference on Signal, Image, Vision and their Applications (SIVA)*, 2018, pp. 1–4, doi: 10.1109/SIVA.2018.8661149.
- [17] R. Gowthaman, “Automatic Identification and Classification of Microaneurysms for Detection of Diabetic Retinopathy,” *Int. J. Res. Eng. Technol.*, vol. 3, no. 02, pp. 464–473, 2014.
- [18] 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*, 2011, pp. 2045–2048.
- [19] F. D. Aferi, T. W. Purboyo, and R. E. Saputra, “Cotton texture segmentation based on image texture analysis using gray level co-occurrence matrix (GLCM) and Euclidean distance,” *Int. J. Appl. Eng. Res.*, vol. 13, no. 1, pp. 449–455, 2018.
- [20] B. Azam *et al.*, “A reliable auto-robust analysis of blood smear images for classification of microcytic hypochromic anemia using gray level matrices and gabor feature bank,” *Entropy*, vol. 22, no. 9, p. 1040, 2020.
- [21] S. Punitha, A. Amuthan, and K. S. Joseph, “Benign and malignant breast cancer segmentation using optimized region growing technique,” *Futur. Comput. Informatics J.*, vol. 3, no. 2, pp. 348–358, 2018, doi: 10.1016/j.fcij.2018.10.005.
- [22] K. Preetha and S. K. Jayanthi, “GLCM and GLRLM based Feature Extraction Technique in Mammogram Images,” *Int. J. Eng. Technol. Website*, vol. 7, pp. 266–270, 2018.
- [23] S. S. Xu, C. Chang, C. Su, and P. Q. Phu, “Classification of Liver Diseases Based on Ultrasound Image Texture Features,” *J. Appl. Sci. MDPI*, vol. 9, pp. 1–25, 2019, doi: 10.3390/app9020342.
- [24] Q. Humaini, “Jaringan Syaraf Tiruan Extreme Learning Machine (ELM) untuk Memprediksi Kondisi Cuaca di Wilayah Malang,” *Univ. Islam Negeri Maulana Malik Ibrahim Malang*, 2015.
- [25] B. Y. Phiadelvira, D. Z. Haq, and D. C. R. Novitasari, “Prediksi Besar Daya Listrik Tenaga Gelombang Laut Metode Oscillating Water Coloumn (PLTGL-OWC) Di Perairan Banyuwangi Menggunakan Extreme Learning Machine (Elm),” *UNNES J. Math.*, 2021.
- [26] I. Ahmad, M. Basher, M. J. Iqbal, and A. Rahim, “Performance comparison of support vector machine, random forest, and extreme learning machine for intrusion detection,” *IEEE access*, vol. 6, pp. 33789–33795, 2018.
- [27] S. Ding, H. Zhao, Y. Zhang, X. Xu, and R. Nie, “Extreme learning machine: algorithm, theory and applications,” *Artif. Intell. Rev.*, vol. 44, no. 1, pp. 103–115, 2015, doi: 10.1007/s10462-013-9405-z.