

Detection of Leukemia by Using K-Means Clustering Algorithm and Support Vector Machine

Marfa Azhari Ahmed 17F17337@mec.edu.om, Puttaswamy Malali Rajegowda
puttaswamy@mec.edu.om, Raza Hasan raza@mec.edu.om, and Robin Zarine
robin@mec.edu.om

Abstract

Leukemia is a type of blood cancer that normally originates in the bone marrow. It causes a relatively large number of abnormal blood cells to be produced. In a normal, healthy state, blood cells originate in the bone marrow as stem cells and later mature to form different types of blood cells (red blood cells, white blood cells, or platelets) and transfer to the bloodstream. As for a person suffering from leukemia, his/her bone marrow begins to produce abnormal white blood cells that enter the bloodstream and begin to compete against the normal healthy blood cells and prevent them from performing their functions properly. This paper aims to detect blood cancer disease (Leukemia) at the earliest stage possible through the use of image processing techniques, k-means clustering algorithm, and support vector machine (SVM). The process will be further supported by processing tools such as MATLAB to ensure the early identification of any leukemia existence, thus enabling early administration of appropriate treatment to mitigate possible pestilent outcomes. The detection through using the images is a cheap, fast, and safe method as there no need for specialized equipment used for lab testing.

Keywords: *Leukemia identification; image processing; clustering algorithm; SVM*

Introduction (Background and Rationale)

Leukemia disease begins in the bone marrow and spreads beyond the control of underdeveloped white blood cells. When leukemia occurs, the body will produce many white cells and becomes unable to prevent and fight infection and various indications, which can even lead to the death of the person with this malignant disease (Webmd,2020). A gene mutation in the DNA of blood-producing cells causes leukemia. The blood cells contain dysfunctional and inactive lymphocytes as a result of this mutation.

The lymphocytes continue to survive and spread, while the normal lymphocytes die. The abnormal lymphocytes build up in the blood and some organs, causing problems, and they may push healthy cells out of the bone, interfering with the production of normal blood cells, and this may result in a variety of symptoms such as weight loss, fever, bleeding, breath shortness, and black skin spots. In addition, according to the symptoms, this disorder has four distinct diagnosis stages (Conrad and Jerry, 2020).

Leukemia is curable if diagnosed and treated early, but detection is a time-consuming operation. This study aims to use an automated machine learning to detect blood cancer (leukemia) disease cells, stage, and recoverability using image pre-processing techniques, segmentation features extraction, and support vector machine (SVM), which will help the doctors and laboratory technicians to examination thousands of patients in a short time and with less investigation.

Problem Statement

The discovery of leukemia can be traced back thousands of years. Scientists have attempted for centuries to come up with new and groundbreaking ways to diagnose this disease, and the outcomes were beneficial, although sluggish and complicated.

Until now, diagnosing leukemia has been a lengthy and difficult procedure for doctors, due to the many blood tests that patients must complete for the doctor to examine and interpret the results of the tests, and if the test results are unclear, the doctor will order further tests, therefore, detection of this disease is one of the complex tasks of the doctor.

In general to solve the complex process of detecting cancer disease a machines learning can be developed in the medical sector using image processing techniques, artificial intelligence, and certain algorithms, to become a useful solution for a variety of diseases, such as our proposed project which is an automated machine that can detect the blood cancer, and increase the quality and accuracy of detecting this disease, which will lead to detect this disease at early stages and save more patients' lives.

Supporting Literature

Beham and Gurulakshmi (2012) proposed an image processing method that can analyze the medical and MRI images to detect the tumor and cancer cells. The proposed method is based on K -means clustering algorithm and morphological image processing which enhance the quality and the performance of the system. Medical Image segmentation deals with the segmentation of tumors in CT and MR images for improved quality in medical diagnosis.

Karolinska Institute (2017) Defines the microscope image as being more accurate than the 2-dimensional image in detecting and diagnosing cancer. The author of this journal article discusses new image processing techniques to detect and diagnose cancer cells by transporting tissues and reproducing a three-dimensional picture.

Kazemi et al (2016) explored automatic algorithms that identify and detect acute methodology leukemia and its subtypes. the proposed algorithm involves lots of techniques such as processing the microscope image, color segmentation strategy to segment the white blood cells from the blood components, discriminative features to extract the texture features from the nucleus, and classified the image by using a support vector machine. the proposed algorithm has achieved an acceptable performance of detecting the acute methodology leukemia and its subtypes.

Kumar et al (2020) have computerized a method for cancer detection to minimizing human involvement. The proposed method was automated image-based acute leukemia detection systems by using incensement and filtering techniques, extract regions using a K-means clustering algorithm. The proposed system has achieved 93% accuracy in the testing phase.

Salem (2014) Segmentation is defined as a lengthy process that necessitates the use of a proper algorithm to achieve a successful outcome. The author of this journal article discusses a new scheme for segmenting blood cancer cells from a microscope image using the k means clustering algorithm, which was tested and evaluated using a blood sample.

Schmidt (2017) developed software that used a modern algorithm to detect blood, thus improving the effective diagnosis of blood cancer since blood vessels are considered to be the key factor in diagnosing blood cancer, this software was implemented using endpoint constants techniques, and it was only tested on animal blood samples. The developed software is used a new approach to detecting blood cancer that focuses on detecting the blood vessel and the blood cancer stage.

Solomon and Breckon (2011) Mentioned that MATLAB may be used to describe various systematic analyses of image processing techniques. The authors have covered a variety of image processing concepts that can be used in

MATLAB and addressed the various types of image processing concepts that can be implemented in MATLAB with appropriate examples, which were extremely helpful in understanding how to develop and implement the image processing technique in MATLAB.

Methods

The following are the proposed methods for implementing automated blood cancer detection (Leukemia) by using microscope images.

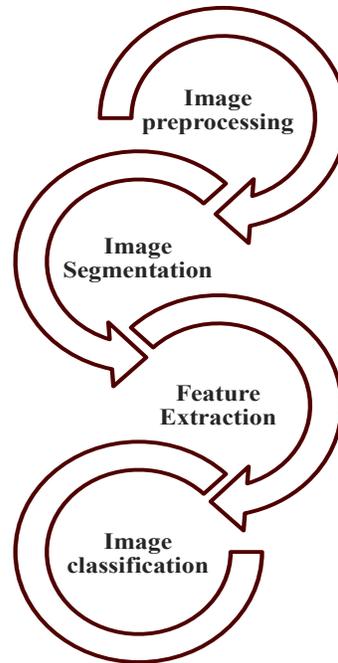


Figure 1. The proposed methods of Leukemia detection.

Image pre-processing

The aim of image pre-processing is to enhance and eliminate noise from the input image (Patel and Mishra, 2015), so the Filtering noises removal method and edge detection using Sobel mask are used in the proposed work to enhance and improve the input image and aid in fast segmentation of WBCs using the k-means clustering algorithm.

Image Segmentation

The WBCs structure, which is irregular cells, is located using image segmentation of microscope blood cell images. Images are segmented because they are separated into a group of pixels. The K means clustering algorithm is used in the proposed work, to aids in the extraction of important characters from microscope images by implementing the repeated segmentation scheme, which explores the centroid of each set in the segment and then re-centroid segment input centered on the closest centroid.

Feature Extraction

This is to ensure that the characteristics of WBCs are collected to determine if the cells are blast or regular cells and that their three characteristics are taken into account when detecting blood cancer cells by using a thresholding-based Gray-level Co-occurrence matrix. The extracted features in this stage are:

- Geometrical parameters such as radius, eccentricity, perimeter, and concavity are used.
- Statistical parameters such as the gradient matrix and histogram are used.
- Textural parameters of WBCs such as homogeneity, energy, and contrast are used.

Image classification

The description of good descriptive features that facilitate their classification is needed to get a suitable identification of blood cancer cells. Using the Support Vector Machine (SVM), which is based on statistical learning theory, the extracted features are analyzed to classify the types of blood cancer cells and to determine the disease level, and recoverability.

Results

The used methodology of implemented machine learning that uses preprocessing techniques, segmentation, feature extraction, and classification with a support vector machine to automatically process a microscope image of blood cells, and identify blood cancer cells and classify them under a certain stage, are very useful method, as it resulted in a clear output that shows the validity of the used method used, and I have discussed these results in the discussion section.

The proposed flowchart is shown the process of the used method in figure 2:

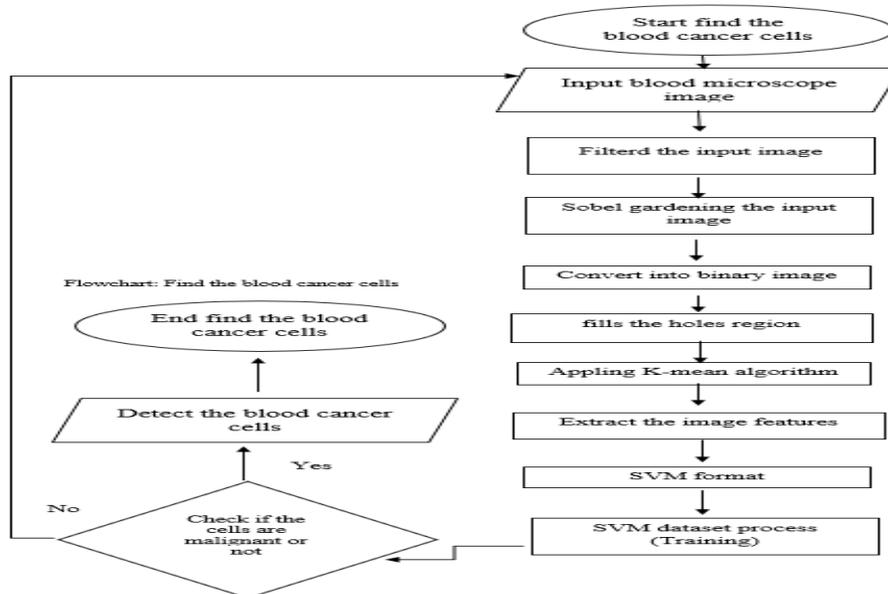


Figure 2. Flowchart of Leukemia detection used methods.

Discussion

The discussion here is about the used methods in detail and how it's detecting the results. Here will start the detection process by using an infected blood cancer microscope dataset figure 3 (Sharma, 2021).

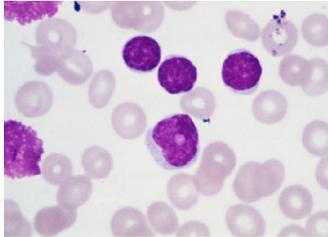


Figure 3. Infected Blood Cancer Microscope Image.

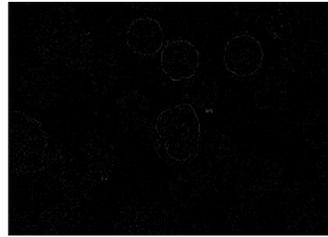


Figure 4. Filtered Image.

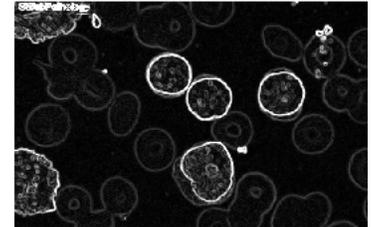


Figure 5. Sobel Gradient Image.

The first step is to apply the image pre-processing techniques to converting the input image figure 3 to grayscale and eliminate the unwanted noises and filtering by using filter noises removal technique figure 4, then edge the detections with the Sobel filter, which considers as a type of gradient-based approach and modifies pixels to estimate gradients in image directions of figure 4 to get the Sobel gradient image figure 5.

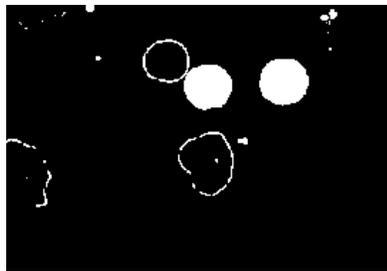


Figure 6. Hole Filled Image.



Figure 7. Binary Detected Cells Image.

The second step is applying the segmentation step, but first, it fills the holes located in figure 5 then it uses the k-means clustering algorithm to extract the input image characteristics and locate the WBCs, which are the abnormal cells. The outcome of the segmentation is transformed into a binary form to evaluate the segmentation scheme in figure 7.

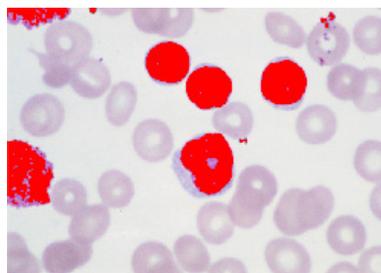


Figure 8. Detected Blood Cancer Cells Image.

The third step is to extract the features of WBCs to ensure and verify whether the extracted features are normal or not, this process is done by using the Thresholding Based Grayscale Level CO Occurrence Matrix. The feature extraction stage extracts the geometrical, statistical, and textural features, and then the classification process is started by using a Support Vector Machine (SVM), by train a microscope dataset in MATLAB (BMC Medical Genomic, 2021) that contains different means such as connectivity and area means, and the data set is randomly divided into train and test to determine if the segmentation's output cells are malignant or not.

The classification is done in the SVM by loading the dataset and using the kernel function to classify the cells, and then the SVM will evaluate the output of the cells based on the testing data.

As showing in figure 8 the blood cancer cells have been detected successfully and marked as a red color, so according to the detected damaged cells the blood cancer is in the second level and it can be recoverable because the percentage of the damaged cells are 26% so it's less than 50%.

The machine learning has been trained and tested by using an infected blood cancer microscope data set (Sharma,2021), the result was so satisfactory because the correct prediction was 980 images from 1000, images so it provides more accuracy of the blood cancer detection around 96% figure 9.

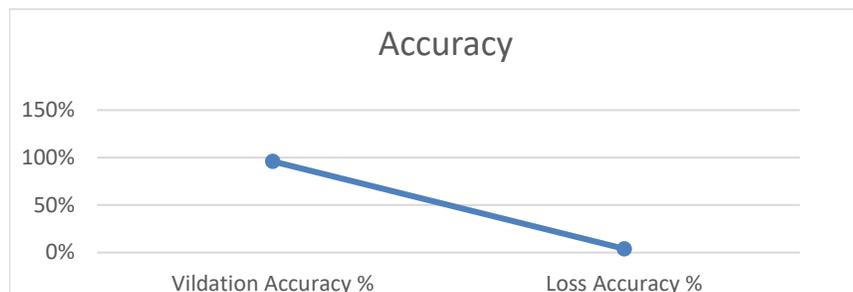


Figure 9. Training and Testing Accuracy.

Conclusion

This paper has proposed an automated machine learning that can detect blood cancer cells, cancer stage, percentages damaged cells, and cell recoverability from a microscope image using the MATLAB tool to reduce detection time and improve accuracy over the manual approach, the use of this method will enhance the traditional process of detection the blood cancer with the less human investigation and better accuracy around 96%.

The use of the proposed method has the potential to save many lives by identifying this disease in its earliest stages when it is easier to recover and helps the laboratory technicians and doctors to detect blood cancer (Leukemia) at early stages without the need for complex and costly blood test thus reducing the patient's expenses and increase the medical field in Sultanate of Oman.

Limitations

The validity of the research and result and its applicability are based on the scope, but it was difficult to use an external application with the MATLAB codes, so instead of that, the MATLAB GUI is used to implement the project GUI.

Acknowledgments

I am very grateful to everyone who contributed and supported me to complete this paper, particularly Allah, and then Mr. Puttaswamy M. R. who supported me to complete my research project and article until the end.

References

1. Amin, M., Kermani, S., Talebi, A., and Oghli, M. (2015) 'Recognition of acute lymphoblastic leukemia cells in microscope images using K-means clustering and support vector machine classifier' *Journal of medical signals and sensors* [online] 5 (1), 49-58. available from <<https://europepmc.org/article/PMC/4335145>> [15 November 2020]
2. Beham, M., and Gurulakshmi, B. (2012) 'Morphological image processing approach on the detection of tumor and cancer cells' *2012 conference on devices, circuits and system* [online] 15 (2), 350-354. available from <<https://ieeexplore.ieee.org/document/6188786>> [1 December 2020]
3. BMC Medical Genomic (2021) *Table 1 Dataset content from PrognosSan* [online] online from <<https://bmcmmedgenomics.biomedcentral.com/articles/10.1186/1755-8794-2-18/tables/1>> [28 April 2021]
4. Conrad, M., and Jerry, R. (2020) *Leukemia* [online] available from <<https://www.medicinenet.com/leukemia/article.htm>> [15 November 2020]
5. Jagadeesh, S., Nagabhooshanam, E., and Venkatachalam, S. (2013) 'Image processing approach to cancer cell prediction in blood samples' [online] 1 (5), 2-7. available from <<https://www.semanticscholar.org/paper/IMAGE-PROCESSING-BASED-APPROACH-TO-CANCER-CELL-IN-Jagadeesh-Nagabhooshanam/22583fde46f9ce31a515ad053ede14cb3618f011#citing-papers>> [10 November 2020]
6. Karolinska Institute (2017) *3D microscopy gives more accurate cancer diagnosis* [online] available from <<https://www.sciencedaily.com/releases/2017/10/171003094825.htm>> [3 December 2020]
7. Kazemi, F., Najafabadi, T., and Araabi, B. (2016) 'Automated Recognition of Acute Myelogenous leukemia in blood microscope image using K-means clustering and support vector machine' *Journal of medical signals and sensors* [online] 6 (3), 183-193. available from <<https://www.ncbi.nlm.nih.gov/pmc/articles/PMC4973462/>> [27 December 2020]
8. Kumar, S., Sumita, M., Pallavi, A., and Pragya, A. (2020) 'Automated detection of acute leukemia using K-mean clustering algorithm' *Department of chemistry, MVPG college* [online] 10 (2), 2-14. available from <<https://arxiv.org/ftp/arxiv/papers/1803/1803.08544.pdf>> [30 November 2020]
9. Mohapatra, S., Patra, D., and Satpathy, S. (2011) 'Automated leukemia detection in blood microscope images using statistical texture analysis' *International conference on communication, computing & security* [online] 1(7), 184-187. available from <https://www.researchgate.net/publication/220846420_Automated_leukemia_detection_in_blood_microscopic_images_using_statistical_texture_analysis> [27 November 2020]
10. Rajegowda, P.M. and Balamurugan, P., 2018. A Neural Network Approach to Identify Hyperspectral Image Content. *International Journal of Electrical and Computer Engineering (IJECE)*, 8(4). DOI: [10.11591/ijece.v8i4.pp2115-2125](https://doi.org/10.11591/ijece.v8i4.pp2115-2125).
11. Patel, N., and Mishra, A. (2015) 'Automated leukemia detection using microscope image' *Provide computer science* [online] 58 (1), 635-642. available from <<https://www.sciencedirect.com/science/article/pii/S1877050915021936>> [20 November 2020]
12. Puttaswamy, M.R. & Balamurugan, P. (2018). Hyperspectral image content identification using kernel based neural network. *Journal of Advanced Research in Dynamical and Control Systems (JARDCS)*. 10. 107-115.
13. Salem, N. (2014) 'Segmentation of white blood cells from microscope images using k-means clustering' *National radio science conference (NRSC)* [online] 5 (1), 14-50. available

from<https://www.researchgate.net/publication/262185226_Segmentation_of_White_Blood_Cells_from_Microscopic_Images_using_K-means_Clustering> [6 December 2020]

14. Sharma, N.(2021) *Leukemia dataset* [online] online from <<https://www.kaggle.com/nikhilsharma00/leukemia-dataset>> [14 April 2021]
15. Solomon, C., and Breckon, T. (2011) *Fundamentals of digital image processing*. United Kingdom: Oxford
16. WebMD (2020) *Types of blood cancer* [online] available from <<https://www.webmd.com/cancer/lymphoma/types-and-differences#1>> [1 November 2020]