

Identification of Cancer Disease Using Image Processing Approaches

Saif Ali, Aneeqa Tanveer, Azhar Hussain, Saif Ur Rehman^{*}

University Institute of Information Technology, PMAS Arid Agriculture University, Rawalpindi, Pakistan

Email address:

Saif.ali3107@gmail.com (S. Ali), Aneeqatanveer@hotmail.com (A. Tanveer), Azhar.uzair2@gmail.com (A. Hussain),

Saif@uaar.edu.pk (S. Ur Rehman)

^{*}Corresponding author

To cite this article:

Saif Ali, Aneeqa Tanveer, Azhar Hussain, Saif ur Rehman. Identification of Cancer Disease Using Image Processing Approaches. *International Journal of Intelligent Information Systems*. Vol. 9, No. 2, 2020, pp. 6-15. doi: 10.11648/j.ijis.20200902.11

Received: May 14, 2020; **Accepted:** June 2, 2020; **Published:** July 4, 2020

Abstract: Cancer, also called malignancy, is an abnormal growth of cells. There are more than 100 types of cancer, including breast cancer, skin cancer, lung cancer, colon cancer, prostate cancer, and lymphoma. Symptoms vary depending on the type. Cancer treatment may include chemotherapy, radiation, and/or surgery. According to American Cancer Society America will be encountering 1,806,950 new cases of cancer in the year 2020 causing 606,520 deaths. Cancer is the leading cause of death in the world. Cancer can be classified into two main categories malignant and benign. Early detection of cancer is the key to the successful treatment of cancer. There are various methodologies for the detection of cancer some include manual procedures, Manual identification is time-consuming and unreliable therefore computer-aided detection came into the research. Computer-aided detection involves image processing for feature extraction and classification techniques for the recognition of cancer type and stages. In this paper, several different algorithms have been discussed such as SVM, KNN, DT, etc. for the classification of the different cancers. This paper also presents a comparative analysis of the researches done in the past.

Keywords: Image Processing, Acute Lymphoblastic Leukemia, ALL, Blood Cancer, Image Segmentation, Performance, Efficiency

1. Introduction

Cancer refers to one of many diseases characterized by the development of abnormal cells that divide uncontrollably and can infiltrate and destroy normal body tissue. Throughout our lives, healthy cells in our bodies divide and replace themselves in a controlled fashion. Cancer starts when a cell is somehow altered so that it multiplies out of control. There are more than 100 types of cancer, including breast cancer, skin cancer, lung cancer, colon cancer, prostate cancer, and lymphoma. Symptoms vary depending on the type.

According to the American Cancer Society, Blood cancer is considered the world's 7th most death-causing cancer. From 2013 to 2017 approximately 15 people out of 100,000 were diagnosed with blood cancer and had a mortality rate of 50 percent. Approximately every 3 minutes one person in the United States (US) is diagnosed with blood cancer. An estimated combined total of 176,200 people in the US is expected to be diagnosed with blood cancer. Furthermore,

they have also estimated to have 60,530 new cases of leukemia and 23,100 death cases in 2020 [1].

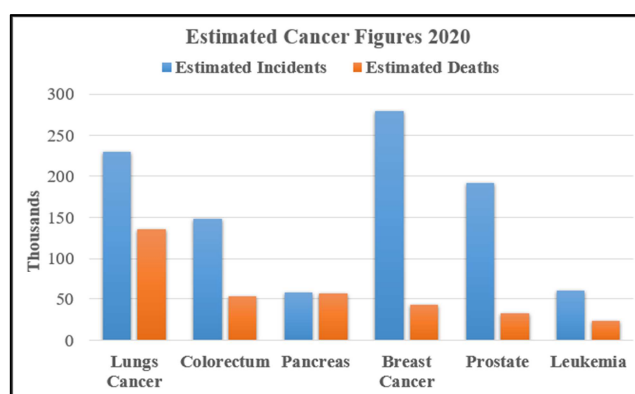


Figure 1. Estimated Incidents and Deaths of Cancer in 2020 [1].

Blood Consists of four major components i.e. Red Blood Cells, White Blood Cells, Platelets, and Plasma. Red cells

contain a special protein called hemoglobin, which helps carry oxygen from the lungs to the rest of the body and then returns carbon dioxide from the body to the lungs so it can be exhaled. White blood cells protect the body from infection. They are much fewer in number than red blood cells, accounting for about 1 percent of your blood. Platelets help the blood clotting process (or coagulation) by gathering at the site of an injury. The main job of the plasma is to transport blood cells throughout your body [2].

Leukemia is a blood cancer that originates in the blood and bone marrow. It occurs when the body creates too many abnormal white blood cells and interferes with the bone marrow's ability to make red blood cells and platelets. Two types of abnormal white blood cells can turn into leukemia: lymphoid cells and myeloid cells. When leukemia caused due to lymphoid cells is called lymphocytic or lymphoblastic leukemia and if it is found in the myeloid cells, it is called myelogenous or myeloid leukemia [3]. The four major types of leukemia are:

1. Acute lymphoblastic leukemia (ALL)
2. Acute myelogenous leukemia (AML)
3. Chronic lymphocytic leukemia (CLL)
4. Chronic myelogenous leukemia (CML)

Acute or chronic which are grouped on the basis of how fast the cells grow. Lymphoid or myeloid is another type of leukemia that is categorized depending on the type of white blood cell that has turned into leukemia [4].

The infected cells can be observed in the microscopic image, manually by a trained expert where the unique features are visually observed and the classification of the type of cancer is done. The variety of features and often unclear images results in missing data which can be a vital indicator to differentiate the type of cancer, therefore identification task becomes difficult. [5].

According to the authors [23]. Many of the diseases can be identified by the blood test of the patient as blood is composed of a significant amount of chemicals and therefore

considered as the river of life. In the case of cancer, the blood test is not enough to identify the cancer type and propagation ratio in the body of the victim.

The biopsy is the recent and one of the most advanced technologies to diagnose a problem or to help determine the best therapy option, especially in cancer. [24, 25]. A biopsy is a sample of tissue taken from the body in order to examine it more closely. A doctor should recommend a biopsy when an initial test suggests an area of tissue in the body isn't normal [26]. Biopsies are most often done to look for cancer. But biopsies can help identify many other conditions.

A: The biopsy instrument is 'primed' prior to insertion under ultrasound guidance. Most disposable instruments offer the option of 1 or 2 cm core lengths. The biopsy needle is slowly advanced to the edge of the target lesion.

B: The central stylet is then slowly advanced through the lesion whilst keeping the remainder of the instrument still. The notch can usually be visualized easily and so the operator can confirm that the target tissue will fall within the biopsy specimen.

C: Once the operator is satisfied with the position of the central stylet/notch, the instrument is 'fired' by further firm forward pressure on the 'plunger', which rapidly advances the outer cutting sheath over the central stylet and samples a core of tissue.

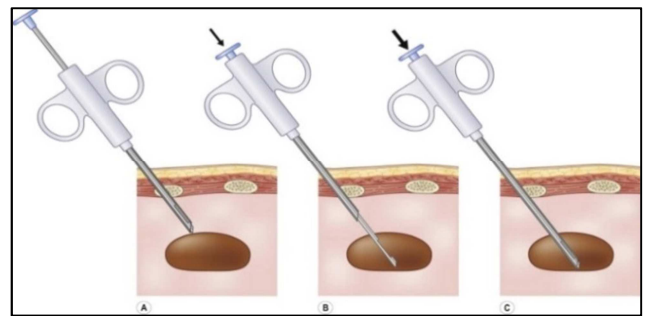


Figure 2. Steps Involved in Biopsy.

Table 1. Various Classifiers for the Computer Aided Detection.

S. No	Author	Year	Methods	Accuracy
1	Janee Alam 1 et al [27]	2018	SVM, GLCM	84.3%
2	Moffy Vas et al [28]	2017	Horlick feature, Artificial Neural Network	92%
3	Suren Makaju [29]	2017	SVM, GLCM	<=100%
4	P. Bhuvaneswari et al [30]	2015	KNN, Gabor filter	90%
5	Manasee Kurkure 1 et al.[31]	2016	Genetic candidate group search algorithm, naive base classifier	82%
6	S. Sasikala et al.[31]	2018	Deep CNN	96%
7	A. Asuntha et al.[33]	2016	SVM, Super pixel segmentation	89.5%

The remainder of this paper is organized as follows. Section II provides a literature review of various techniques used in computer-aided detection of cancer using image recognition. Section III comprises of the generic steps involved in the identification of the cancer detection which are Microscopic Image Acquisition, Image Pre-Processing, Image Segmentation, Feature Extraction, Feature Selection, Classification respectively. Further, in Section IV Comparative analysis of the techniques used are under the observation. The Last Section concludes the work is done in

this study, limitations and future direction are discussed.

2. Literature Survey

2.1. Image Processing-based Approach to Cancer Cell Prediction in Blood Samples

Symptomatic radiography assigns the technological parts of medicative pictures and specifically in acquiring therapeutic images. In [6], authors presented the pre-processing strategies

for leukemia injected cells where the final aim is to generate the elements which describe types of leukemia. The undertaken issues contain the cell segmentation [6] by using the watershed change, determination of distinct cells, and texture quality, statistical and geometrical examination of the cells

2.2. An Overview of Melanoma Detection in Dermoscopy Images Using Image Processing and Machine Learning

Mishra [7] has used image processing techniques to identify melanoma from the microscopic images of dermoscopy. Melanoma recognition using dermo copy images has the most probable for the distraction of the current clinical standard is a fast, accurate and cost-effective on-the-spot technology. Dermoscopy (also known as dermatoscopy or epiluminescence microscopy) is a method of acquiring a magnified and illuminated image of a region of skin for increased clarity of the spots on the skin. The imaging instrument used for this purpose is called a dermatoscope. Dermatoscopes are of two types: contact, using a layer of gel/oil applied between skin and dermatoscopy, and non-contact, with no skin contact and no fluid. Non-contact images, and some contact images, use cross-polarized light from the dermatoscope to acquire the image. Dermoscopy images, because of their illumination and magnification, are widely used in the analysis and examination of skin lesions. Schematic steps involved in the identification of melanoma using dermoscopy images:

Lesion segmentation: Segmentation of the lesion means separating that region (lesion) from the normal skin region

Feature segmentation: In this segment, the region is observed closely and in addition to their presence, the distribution of a feature in the lesion area provides further diagnostic information.

Feature Generation and Classification: Predicting a lesion to be benign or malignant is a binary classification problem. In some cases, it is also important to examine the attributes of the surrounding regions for proper discrimination of melanoma.

Various Classifiers such as k-NN, SVM, ANNs can be explored, the evaluation of classifier results are based on overall accuracy, sensitivity, and specificity of the system [7].

2.3. Breast Cancer Detection Using Image Processing Techniques

In another study [8], a novel system is proposed for the early detection of breast cancer as it is the most common among women and the second most death-causing cancer. The paper proposes the following framework for the identification of the cancer. Mammography is currently the best method for detecting breast cancer at its early stage. The problem with mammography images is they are complex. Detecting macrocalcification in dense breast tissue can be a difficult task as both tend to depict white pixels on the mammogram. The number of false-positive cases on dense breast tissue is higher. Indicators of cancer symptoms are generally, masses and

microcalcifications. Detecting masses are a more challenging task than the detection of microcalcifications. As their size and shape vary in large variation and they often exhibit poor image contrast. Thus, image processing and feature extraction techniques are used to assist radiologists in detecting tumors. The following steps are to be followed for the proposed system.

Image Processing: The general methods for image preprocessing are divided into various branches such as image enhancement, noise removal, image smoothing, edge detection and enhancement of contrast.

Thresholding Techniques: Thresholding is an old, simple and popular technique for image segmentation.

Global Thresholding (GT): is one of the most common and most used techniques in image segmentation. As masses usually have greater intensity than the surrounding tissue. A global thresholding value can be found based on the histogram of the image. On the histogram, the regions with an abnormality impose extra peaks while a healthy region has only a single peak.

Image Segmentation: Partitioning an image into regions such that each region is homogeneous with respect to one or more properties (such as brightness, color, texture, reflectivity, etc.). Common image segmentation methods are thresholding, edge-based segmentation, region-based segmentation, clustering, classifier based segmentation, and deformable model-based segmentation.

Feature Extraction and Selection: Feature extraction is a very important process for the overall system performance in the classification of micro-calcifications. The features extracted are distinguished according to the method of extraction and the image characteristics. The features which are implemented here are texture features and statistical measures like Mean, Standard deviation, Variance, Smoothness, Skewness, Uniformity, Entropy, and kurtosis.

Classification and Evaluation: Evaluation is done based on the acquired features and these features are compared to the respective reference to draw a final conclusion.

Neural Networks: All these values of the texture features are stored and passed through the Neural Network. Back Propagation algorithm can be used to find a pattern within the datasets to automatically finding cancer. A backpropagation algorithm can be designed to self-learn and adjust the weight accordingly. As a greater number of data are entered into Neural Network the better the pattern recognition and accuracy.

2.4. Digital Image Analysis in Breast Pathology—from Image Processing Techniques to Artificial Intelligence

In [9], a deep learning paradigm is suggested for the early detection of Breast cancer. Authors claimed that the breast cancer is the most common malignant disease in women worldwide. Diagnosis by histopathology has proven to be instrumental to guide breast cancer treatment, but new challenges have emerged as our increasing.

The complexity of and demand for accuracy in histopathologic breast cancer diagnosis is increasing.

However, the lack of pathologists is an evident issue in most parts of the world. As patient demand for personalized breast cancer therapy grows, the world faces an urgent need for more precise biomarker assessment and more accurate histopathologic breast cancer diagnosis to make better therapy decisions. The digitization of pathology data has opened the door to faster, more reproducible, and more precise diagnoses through computerized image analysis.

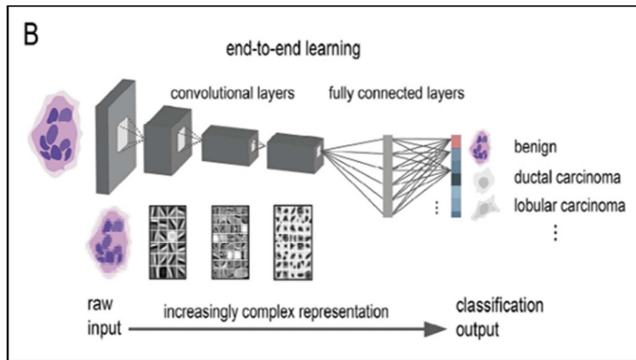


Figure 3. Fig From [9]. Deep Learning paradigm for the detection of Breast Cancer.

In contrast, deep learning is an end-to-end approach to learning that takes raw images as input and directly learns a model to produce the desired output. Deep learning uses biologically inspired networks to represent data through multiple levels of simple but nonlinear modules that transform the previous representation into a higher, slightly more abstract representation. The compositional nature of the architecture allows deep networks to form highly complex and nonlinear representations as each layer forms a more abstracted representation than the last. The result is a rich representation that provides unprecedented discriminatory power.

2.5. Computer-Aided Acute Lymphoblastic Leukemia Diagnosis System Based on Image Analysis

The main objective of the study presented in [10] is to identify the lymphocyte by segmenting the microscopic images then diagnose (classify) each segmented cell to be normal or affected. Computer-Aided Acute Lymphoblastic Leukemia (ALL) diagnosis system based on image analysis. Leukemia is a kind of cancer that basically begins in the bone marrow. Children under 5 years and older people over 50 years are at higher risk of acute lymphoblastic leukemia, also, it can be fatal if it is not treated earlier as it is rapidly spread into some vital organs and the bloodstream too. Acute lymphoblastic leukemia can be diagnosed by the morphological identification of lymphoblasts by microscopy. Blood samples can be observed and diagnosed with different diseases by doctors. Any human-based diagnosis suffers from nonstandard precision as it basically depends on the doctor's skill; also it is unreliable from a statistical point of view. Automated diagnosing systems are more accurate and not temperamental like human-based systems. Also, they are statistically reliable and can be generalized. So, the white blood cell (WBC) affected by acute lymphoblastic leukemia

will be counted and classified. The proposed computer-aided acute lymphoblastic leukemia diagnosis system aims to optimally select the most powerful features that can be used in the lymphoblastic leukemia diagnosis system. The proposed lymphoblastic leukemia diagnosis system consists of three basic phases [10]: Image segmentation, feature extraction, and classification.

2.5.1. Cell Segmentation Phase

Unlike many methods in the literature, the proposed system detects the nuclei and the entire membrane at the same time. The images are in RGB color space which is difficult to be segmented. So, the images were converted to CMYK color space.

2.5.2. Feature Extraction Phase

Feature extraction in image processing is a technique of transforming the input data into the set of features. Three types of features were extracted from the segmented cells including shape features, color features, and texture features.

2.5.3. Feature Normalization

To narrow down the gap between the highest and the lowest value of extracted features and to improve the classification results. Three different normalization techniques were applied; they include grey-scaling, min-max, and Z-score techniques.

Grey-Scaling: It is an image normalization technique used to convert a matrix to a greyscale image. This can be performed by scaling the entire image to the range of brightness values from 0 to 1. It works by normalizing each individual columns or rows to a range of brightness values from 0 to 1.

Min-Max: Data are scaled to a fixed range usually 0–1. The cost of having this bounded range in contrast to standardization is that it will be ended up with smaller standard deviations, which can suppress the effect of outliers.

Z-Score: In this normalization method, the mean and the standard deviation of each feature are calculated. Next, the mean was subtracted from each feature. Finally, the product values were divided by the standard deviation.

2.6. Detection of Blood Cancer in Microscopic Images of Human Blood Samples a Review

In [3], the authors have discussed the identification of leukemia in the early phase, provide the appropriate treatment for the patient. For this problem, the system sets forth the solution that signifies the leucocytes in the blood image and then selects the lymphocytes cells. It assesses the morphological index from those cells and finally, it allocates the existence of leukemia. In their study an image processing techniques have been used to count the number of blood cells in the biomedical image. The original image is converted to a grayscale image for which a threshold value of intensity is set in order to differentiate WBC from RBC. The results acquired using the thresholding technique shows that the ratio of RBC and WBC for a normal image to the abnormal image has a different range of ratio. For normal images, the ratio is 0 to 0.1 whereas for abnormal images its ratio range

is 0.2 to 2.5 for ALL and 0 to 14 for AML. Furthermore, various image processing techniques have been used to detect blood cancer in biomedical images of human blood samples. Like the thresholding technique. In addition to this, authors used an image processing techniques to count the number of blood cells in the biomedical image. With this counted value of blood cells, the ratio of blood cells for leukemia is

calculated. The Methodology used for [3] is medical image recognition.

1. Microscopic Image Acquisition
2. Image Enhancement (Preprocessing)
3. Image Segmentation
4. Image Feature Extraction
5. Image Classification

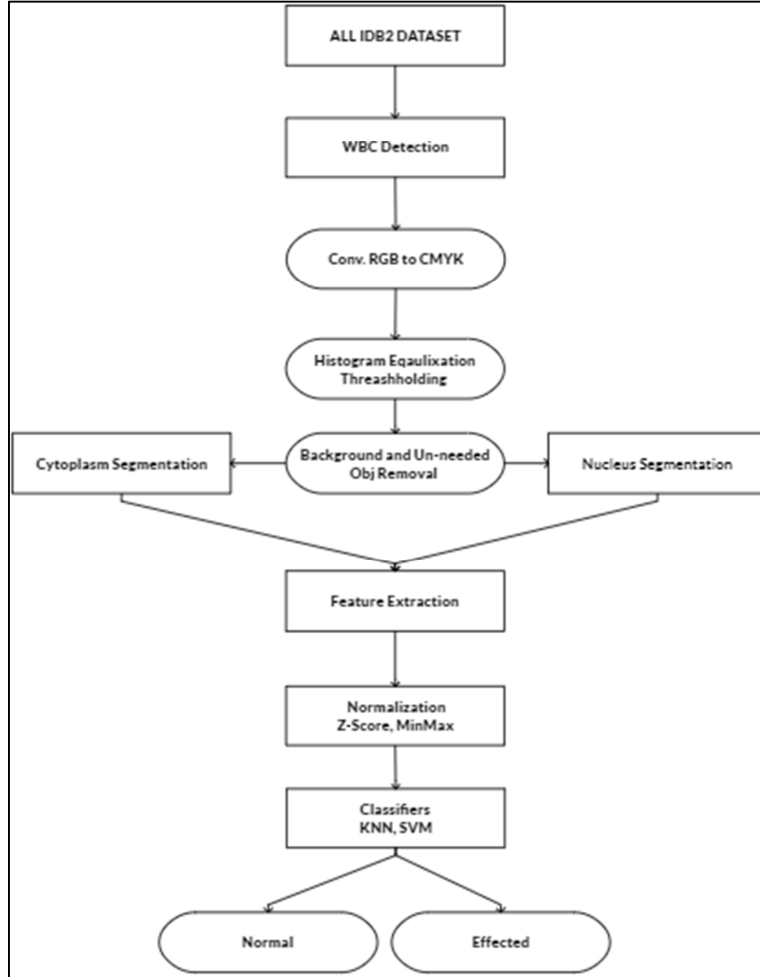


Figure 4. From [10]. The proposed framework for computer-aided detection of Acute Leukemia.

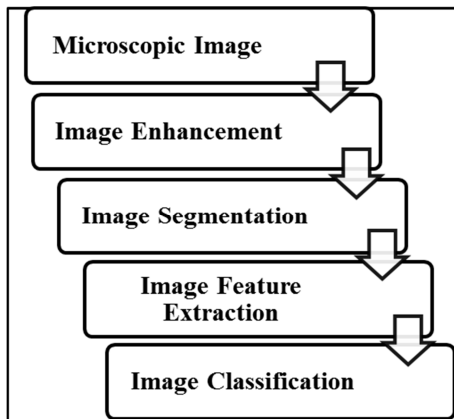


Figure 5. Fig From [3]. A proposed framework for computer-aided detection of Acute Leukemia.

2.7. Automatic detection of Acute Lymphoblastic Leukemia Using Image Processing

The study proposed in [11] implemented a fully automated algorithm by using image processing to aid in the detection of acute lymphoblastic leukemia in identifying and counting the infected white blood cells present in the human blood sample. The early detection and treatment of blood cancer for recovery. Their work presents a method to automatically identify and count the lymphoblast cells in each blood sample, so as to eliminate human errors and most importantly facilitate earlier detection of acute lymphoblastic leukemia. MATLAB is used, with the image processing toolbox, for implementation. 108 image samples are taken from healthy and infected patients, with optical laboratory microscope and Canon PowerShot G5 camera. The images are in .JPG format with 24-bit color depth and a resolution of 2592x1944. The implementation occurs in

five stages; (1)- The first stage identifies the lymphoblasts based on its physical characteristics and separates it from the rest of the blood sample.; (2)- The second stage is the separation of grouped and individual lymphoblasts; (3)- The third stage involves the separation of clustered lymphoblasts by application of the distance transform of watershed segmentation; (4)- The fourth phase involves removing abnormal and unwanted cell components by shape control; (5)- Fifth and the final stage deals with the counting of detected lymphoblasts and calculating the accuracy of the method.

Background removal using the Zack algorithm: The background is removed effectively using the Zack algorithm. The Zack algorithm or triangular method is used to find the thresholding value required for image segmentation.

Separation of grouped and ungrouped lymphoblasts: The next stage deals with separating grouped and ungrouped lymphoblasts. The separation is done based on the roundness ratio. The roundness ratio is the ratio of the square of perimeter and four times pi times its area.

Separation of grouped lymphoblasts: Separation of grouped cells offers great challenges due to the irregular shape of each cell. The separation can be done by the application of watershed segmentation.

Removal of abnormal cell components: Removal of unwanted cell structures and other elements is performed as explained in previous sections. The separated grouped lymphoblasts and ungrouped lymphoblasts are combined and the number of cells in that blood sample is counted.

Table 2. Comparative analysis of histopathology techniques.

Authors Who Proposed The Algorithm	Algorithm For Analysis	Accuracy Result	Difficulties With This Method
Basavanhally et al. [14]	Active Contour model based on Color Gradient with Hierarchical Normalized cut	89% segmentation accuracy	The false-positive rate is a little high
Dundar et al. [15]	Gaussian Mixture Model-based spatial information and Expectation-Maximization (EM) Algorithm	The overall accuracy of 87.9%	Requires parameter optimization
Tosun and GunduzDemir [16]	Graph Run-length matrices	99% accuracy for segmentation	High computational complexity
Veta et al. [17]	Marker-controlled watershed algorithm and Fast radial symmetry transform	Accuracy of 81.2%	Not suitable for specimens containing larger numbers of nuclei with excessive overlapping and touching of cells
Jain et al. [18]	Active Contour Model (Both region-based and boundary-based) with General Classifier Neural Network (GCNN)	83.47% accuracy	ACM fails in images containing high noise and excessive overlapping nuclei
Karsnas [19]	Intuitive Segmentation using vectorial data	Not mentioned	The region keeps growing without spatial constraint

2.8. Early Skin Cancer Detection Using Computer-Aided Diagnosis Techniques

The key contribution of [20] is the comparative study done between color constancy and skin lesion analysis for early skin cancer detection on an EDRA database. Melanoma is a type of cancer that progresses from the pigment-containing cells known as melanocytes. There are two approaches usually used for early skin cancer detection include color constancy approach and skin lesion analysis. The objective is to hold comparison to evaluate the best approach among these two so that the more accurate results are obtained, and the better treatment is set for the cure of cancer. Researchers have suggested that extensive training is required for a non-invasive systems approach in detecting and subsequently treating melanoma. Due to the wrong treatment of melanoma, clinicians are discouraged from utilizing manual techniques. Inaccurate detection of melanoma may lead to erroneous treatment; hence doctors should use standardized and automated system-based methods. In addition to this, automated system-based methods are found to be effective for early detection of melanoma. The modules of a portable real-time non-invasive skin lesion examination system are used for the initial detection of skin cancer. The framework used in the paper is called a Bag of Features BoF is used to classify dermoscopic images in a single stretch (pass). Comparative analysis of various

techniques proposed for early skin cancer detection using color constancy is provided and comparative analysis of various techniques proposed for early skin cancer detection using skin lesion analysis is provided. The performance of skin cancer detection using skin color constancy is evaluated considering Sensitivity (SE), specificity (SP), and accuracy (ACC). SE corresponds to the percentage of melanomas that are correctly classified. SP is the percentage of correctly classified benign lesions and AAC is defined as follows:

$$ACC = \frac{SE+SP}{2} \quad (1)$$

K-fold cross-validation is used to optimize the hyper-parameters of the color constancy system. The skin lesion analysis system was deployed in mobile phones using an android application. In this method, the user needs to provide information such as age, UV exposure and estimated skin lesion age. The skin images are either browsed or they can be acquired with a camera. The app will segment the images, features are extracted, and segmented images are presented to the user. In this method, no standard databases are used. Authors use 3000 skin lesion image sets of manually classified images. The dataset contained about 800 images with melanoma, 600 with dysplastic nevus and the rest 1600 images with benign nevi. Out of 11 methods, applied SVM outperforms with an accuracy rate of 77.06% with a 0.3911

Root Mean Square Error, 1.0 True Positive Rates, and 0.0 False Positive Rates.

2.9. Image Processing Based Leukemia Cancer Cell

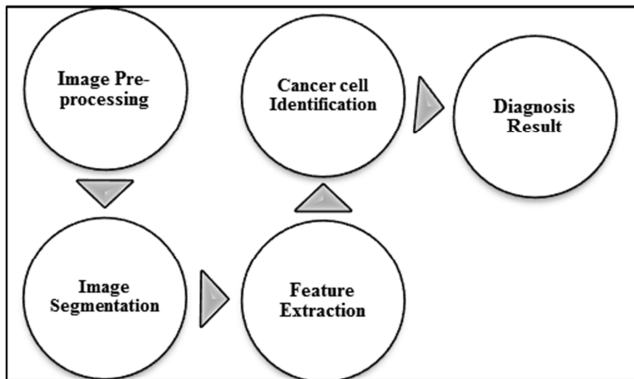


Figure 6. Training and Testing Module of Proposed system in [21].

The objective of the paper [21] is to generate an element which describes, whether the cell is cancerous or not and also identifies the type of leukemia, where mainly leukemia are of four types, they are acute myeloid leukemia, acute lymphocytic leukemia, chronic myeloid leukemia, and chronic lymphocytic leukemia and to overcome drawbacks such as time-consuming analysis, less accuracy and depending on operators' skills. Microscopic pictures are reviewed visually by hematologists and the procedure is tedious and time taking which causes late detection. Therefore, an automatic image handling framework is required that can overcome related limitations in the visual investigation, which provide early detection of disease and type of cancer. The proposed system is providing an environment that has a pre-processing strategy for target cells to tell whether the cell is infected or not. K-means algorithm is used for segmentation. The proposed system reduces human error, time- consuming

analysis and reducing the cost of the treatment and helping the pathologist to suggest effective medication. The proposed system has two parts, training and testing. Both parts undergo the following steps: Image acquisition is the initial step, collecting images of the blood from a microscope with proper magnification from any of the hospitals. The second step is image preprocessing, where the following steps are followed: Initially, color conversion takes place, a color image is converted to a grayscale. Followed by filtering the image, removal of noise from the image and finally histogram equalization is done to increase the quality of an image in terms of contrast. The third step is segmentation using k-mean clustering, and the nucleus is concentrated for the detection process. Segmentation is followed by feature extraction where features of the nucleus are extracted using GLCM and GLDM. In the training part features of the pure cancer cell are stored in the knowledge base. In the testing part, the cell which needs to be tested is, taken as input.

2.10. Lung Cancer Detection Using Medical Images Through Image Processing

The aim of [22] is to design a system which can take any one of the two images, Computed Tomography (CT), Magnetic Resonance Imaging (MRI) scan image and produces an output. Lung cancer is a type of tumor that gets bigger in size and enters other organs of the body. There are various ways to detect lung cancer by using Computed Tomography (CT) scan image, Magnetic Resonance Imaging (MRI) scan image, Ultrasound image. Image processing of the necessary part of the lungs is used for early diagnosis. For this, a system is developed which will help the doctors to easily detect cancer in the lungs from any one of the two images given as input and gives proper analysis. Moreover, CT scan image and MRI scan image are used for experimentations and analysis.

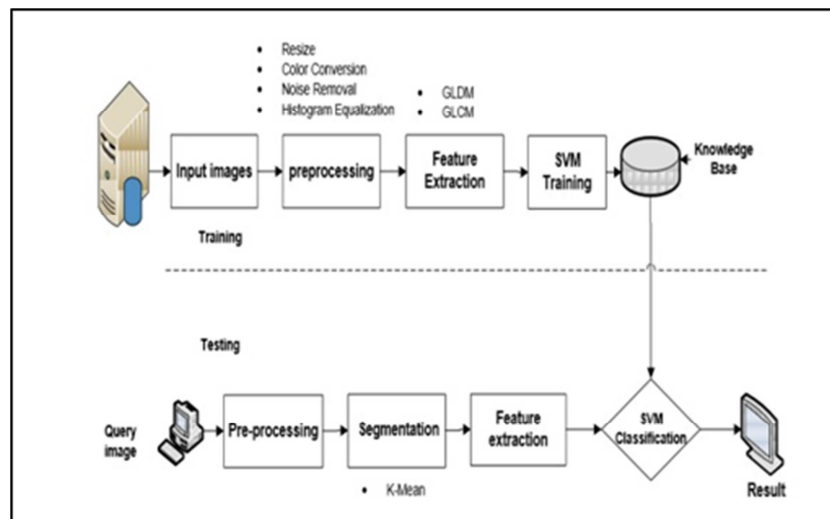


Figure 7. Process Diagram for detection of Cancer cells from MRI AND CT Images [22].

Image pre-processing: Image pre-processing is used to reduce noise and prepare the images for further steps such as segmentation.

Image enhancement: The various image enhancement techniques can be categorized as spatial domain methods and Frequency domain methods. Different image enhancement

techniques are used for all the different images. This includes smoothing of image and, removal of noises, blurring, etc. Gabor filter was found to be suitable for both the CT and MRI images. The filtering of the image proves to be useful for further steps.

Feature Extraction: The Image feature extraction stage is an essential step that represents the final output and determines the normality and abnormality of an image using algorithms and techniques

3. Methodology

After studying the literature, the following general method for automatic blood cancer detection is proposed.

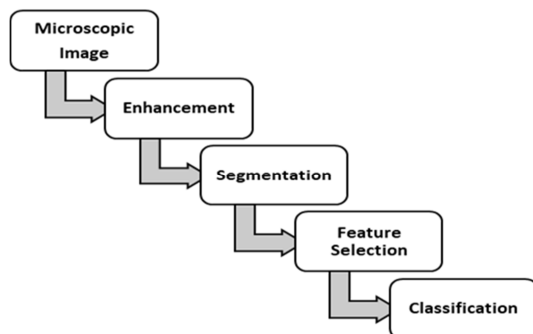


Figure 8. General Paradigm for Automatic Cancer Detection.

3.1. Microscopic Images

Cancer infected blood cell images are collected from the authorized laboratory or from any government hospital in order to carry out further processing.

3.2. Enhancement

Images may contain some artifacts initially, so there is a need to enhance the captured images. Most images contain some sort of noise, so before proceeding further these artifacts should be removed with the help of image enhancement techniques like to remove noise use various kinds of operators i.e. Prewitt and Sobel, canny, etc.

3.3. Segmentation

Segmentation is a process of partitioning an image into subparts, so that proper each and every area is scanned properly. Microscopic images consist of red blood cells, white blood cells, and platelets. But in order to detect the presence of blood cancer, only the number of white blood cells are needed. So, with the help of the segmentation process, the separation of the white blood cells from red blood cells and platelets is achieved. Various techniques used for segmentation are region-based segmentation, k-means Zack algorithm, morphological operation, gradient magnitude, and watershed transform, etc.

3.4. Feature Selection

In this phase, the main focus is to extract some of the features from the processed image. Feature extraction is the

process of converting the image into data so that these values can be compared with the standard values and finally cancerous and noncancerous cells can be separated from the data. Some of the features which are necessary to be calculated are listed below.

3.4.1. Color Features

The mean color values of the grey images are acquired.

3.4.2. Geometric Features

The perimeter, radius, area, rectangularity, compactness, convexity, concavity, symmetry, elongation, eccentricity, solidity is obtained.

Radius: measured by averaging the length of the radial line segments defined by the centroid and border points.

Perimeter: the total distance between consecutive points of the border

Area: the number of pixels on the interior of the cell, defined separately for the nuclei and for the whole cell; as the features, let assume the area of the nucleus and the ratio of the areas of the nucleus and the whole cell.

$$\text{Compactness} = \left[\frac{\text{perimeter}}{\text{area}} \right]^2 \quad (2)$$

Concavity: the severity of concavities in a cell

Concavity points: the number of concavities, irrespective of their amplitudes

Symmetry: the difference between lines perpendicular to the major axis to the cell boundary in both directions Major and minor axis lengths

3.4.3. Texture Features

The entropy, energy, homogeneity, correlation is obtained.

3.4.4. Statistical Features

The skewness, mean, variance and gradient matrix are obtained.

3.5. Classifier

In this final phase, the extracted features are used to provide the final answer. All features extracted are listed in the different columns with their values. When an image is provided as an input to the proposed system than the first step is to calculate the feature values. The values of the test image features are checked with the previously calculated values Based on the values of the input image the classifier classifies that test image into either infected or not infected class.

4. Conclusion

Different categories of classification techniques and computing methodologies applied for cancer diagnosis using images are presented in this paper. The categorization of these techniques with their advantages and disadvantages are discussed.

This work agrees with other researcher's findings such that texture feature analysis is dependent on the quality of the images of cancerous cells, and statistics modeling could be

inaccurate in some specific situations. Each method own some sort of advantages and disadvantages. Diagnostic results of cancer using computer-aided detection methods are depends on the type of images and imaging techniques. The images play foremost roles in determining the results of a cancer diagnosis. In this work, we reviewed various statistical and machine learning methodologies that perform analysis with a texture feature of images and different data pre-processing techniques. We presented a comparative analysis of various techniques based on their performance. This paper conclude based on our review results that a suitable selection of single or combination of machine learning / soft computing algorithms depends upon the data set, capable to yield results with the accuracy of more or equal to 95% on the earlier detection of cancer.

References

- [1] <https://cancerstatisticscenter.cancer.org/#!/cancer-site/Leukemia> [Accessed 5 Feb 2020].
- [2] <https://www.hematology.org/Patients/Basics/#> [Accessed 5 Feb 2020].
- [3] M. Saritha, B. Prakash, K. Sukesh and B. Shrinivas. (2016). Detection of blood cancer in microscopic images of human blood samples: A review. 596-600. 10.1109/ICEEOT.2016.7754751.
- [4] R., Adollah, M. Y., Mashor, N. F. M, Nasir, H., Rosline, H., Mahsin, H., Adilah. (2008). Blood Cell Image Segmentation: A Review. Biomed2008, Proceedings 21, 2008, pp. 141-144.
- [5] Y. Chandni and Z. Shrutika. (2018). Automatic Blood Cancer Detection Using Image Processing. International Journal of Recent Trends in Engineering & Research (IJRTER), vol. 04, no. 03, March 2018, ISSN 2455-1457.
- [6] S. Jagadeesh, E. Nagabhooshanam and S. Venkatachalam. (2013). Image processing based approach to cancer cell prediction in blood samples. International Journal of Technology and Engineering Sciences 1.1 (2013): 1-10.
- [7] Mishra, Nabin K., and Mehmed Emre Celebi. (2016). An Overview of Melanoma Detection in Dermoscopy Images Using Image Processing and Machine Learning. ArXiv abs/1601.07843 (2016).
- [8] Sahni P., Mittal N. (2019) Breast Cancer Detection Using Image Processing Techniques. In: Kumar M., Pandey R., Kumar V. (eds) Advances in Interdisciplinary Engineering. Lecture Notes in Mechanical Engineering. Springer, Singapore.
- [9] R. Stephanie, A. Hossein, S. Kevin and H. Johan (2017). Digital image analysis in breast pathology-from image processing techniques to artificial intelligence. Transl Res. 2017 Nov 7 Published online 2017 Nov 7. doi: 10.1016/j.trsl.2017.10.010.
- [10] Abdeldaim A. M., Sahlol A. T., Elhoseny M., Hassanien A. E. (2018) Computer-Aided Acute Lymphoblastic Leukemia Diagnosis System Based on Image Analysis. In: Hassanien A., Oliva D. (eds) Advances in Soft Computing and Machine Learning in Image Processing. Studies in Computational Intelligence, vol 730. Springer, Cham.
- [11] V. Shankar, M. M. Deshpande, N. Chaitra, and S. Aditi, "Automatic detection of acute lymphoblastic leukemia using image processing," 2016 IEEE International Conference on Advances in Computer Applications (ICACA), Coimbatore, 2016, pp. 186-189. doi: 10.1109/ICACA.2016.7887948.
- [12] Sarkar, S., & Das, S. (2016). A Review of Imaging Methods for Prostate Cancer Detection: Supplementary Issue: Image and Video Acquisition and Processing for Clinical Applications. Biomedical Engineering and Computational Biology. <https://doi.org/10.4137/BECB.S34255>.
- [13] M a, Aswathy & Mohan, Jagannath. (2016). Detection of Breast Cancer on Digital Histopathology Images: Present Status and Future Possibilities. Informatics in Medicine Unlocked. 8. 10.1016/j.imu.2016.11.001.
- [14] Basavanthally A, Yu E, Xu J, Ganesan S, Feldman M, Tomaszewski J, Madabhushi A. Incorporating domain knowledge for tubule detection in breast histopathology using O'Callaghan neighborhoods. Proc SPIE 2011; 7963: 796310–25.
- [15] Dundar M, Badve S, Bilgin G, Raykar VC, Jain RK, Sertel O, et al. Computerized classification of intraductal breast lesions using histopathological images. IEEE Trans Biomed Eng 2011; 58 (7): 1977–84.
- [16] A. B. Tosun and C. Gunduz-Demir, C. Graph run-length matrices for histopathological image segmentation, IEEE Transactions on Medical Imaging 30 (3), 2011, 732-566.
- [17] Veta M, van Diest PJ, Kornegoor R, Huisman A, Viergever MA, Pluim JPW. Automatic nuclei segmentation in H&E stained breast cancer histopathology images. PloS ONE 2013; 8 (7): e70221.
- [18] Jain A, Atey S, Vinayak S, Srivastava V. Cancerous cell detection using histopathological image analysis. Int J Innov Res Comput Commun Engng 2014; 2 (12): 7419–26.
- [19] Karsnas A. Image analysis methods and tools for digital histopathology applications relevant to breast cancer diagnosis. Digit Compr Summ Upps Diss Fac Sci Technol 2014: 1128.
- [20] Fernandes, Steven & Chakraborty, Baisakhi & Gurupur, Varadraj & Prabhu, Ananth. (2016). Early Skin Cancer Detection Using Computer-Aided Diagnosis Techniques. Journal of Integrated Design and Process Science. 20. 33-43. 10.3233/jid-2016-0002.
- [21] Rejintal, Ashwini & Aswini, N. (2016). Image processing based leukemia cancer cell detection. 471-474. 10.1109/RTEICT.2016.7807865.
- [22] Neelima Singh, A. Asuntha (2016). L. Lung cancer detection using medical images through image Processing, Journal of Chemical and Pharmaceutical Sciences, pp 1558-1561
- [23] Tamilmani, G. & Sivakumari, S. (2017). A survey on various data mining methods for detecting cancer cells. 242-245. 10.1109/ICSTM.2017.8089160.
- [24] Cinalli, Giuseppe & Onorini, Nicola. (2020). Biopsy of the Tumor. 10.1007/978-3-030-21299-5_4.
- [25] Candotto, Valentina & Pezzetti, F & Scarano, Antonio & Agazzi, A & Spadari, F & Palmieri, A. (2019). Liquid biopsy. Journal of biological regulators and homeostatic agents. 33.

- [26] [Online] <https://www.webmd.com/cancer/what-is-a-biopsy#1> [Accessed 12 Feb 2020].
- [27] J. Alam, S. Alam, and A. Hossan, "Multi-Stage Lung Cancer Detection and Prediction Using Multi-class SVM Classifie," Int. Conf. Comput. Commun. Chem. Mater. Electron. Eng. IC4ME2 2018, pp. 1–4, 2018.
- [28] M. Vas and A. Dessai, "Lung cancer detection system using lung CT image processing," 2017 Int. Conf. Comput. Commun. Control Autom., pp. 1–5, 2017.
- [29] S. Makaju, P. W. C. Prasad, A. Alsadoon, A. K. Singh, and A. Elchouemi, "Lung Cancer Detection using CT Scan Images," Procedia Comput. Sci., vol. 125, no. 2009, pp. 107–114, 2018.
- [30] P. Bhuvaneswari and A. B. Therese, "Detection of Cancer in Lung with K-NN Classification Using Genetic Algorithm," Procedia Mater. Sci., vol. 10, no. Cnt 2014, pp. 433–440, 2015.
- [31] M. Kurkure and A. Thakare, "Classification of Stages of Lung Cancer using Genetic Candidate Group Search Approach," IOSR J. Comput. Eng., vol. 18, no. 05, pp. 07–13, 2016.
- [32] R. Tekade and K. Rajeswari, "Lung Cancer Detection and Classification Using Deep Learning," Proc. - 2018 4th Int. Conf. Comput. Commun. Control Autom. ICCUBE2018, no. 2, pp. 259–262, 2018.
- [33] A. Asuntha, A. Brindha, S. Indirani, and A. Srinivasan, "Lung cancer detection using SVM algorithm and optimization techniques," J. Chem. Pharm. Sci., vol. 9, no. 4, pp. 3198–3203, 2016.