Leukamania Disease Detection using Image Processing

Kannaiah Venkateswaran Sivani

Department of Computer Science, Sri Venkateswara University, Tirupati

***Abstract*—** *In our article, the paper proposed the automated leukemia detection using Image Processing. In a manual method trained physician count WBC to detect leukemia from the images taken from the microscope. This manual counting process is time taking and not that much accurate because it completely depends on the physician's skill. To overcome these drawbacks an automated technique of detecting leukemia is developed. After that an automated counting algorithm is used to count WBC to detect leukemia.Some features like cA, cH, cV, cD, skewness, energy, entropy, homogeneity, standard deviation etc. are extracted and calculated.*

I. INTRODUCTION

Leukemia is generated from the bone marrow. It can be cause of death if treatment is not started at right time. A thin material is situated inside each bone which is known as bone marrow. Every human body has mainly three type of blood cells:-RBC (red blood cell), WBC (white blood cell),PLT(platelets).In this paper main point of concern is to detect leukemia. So we are only concentrating on the count of WBC (leucocytes). Myeloid stem cell emerges into myeloid blast. Lymphoid stem cell also enduces lymphoid blast which will generate only the white blood cell(WBC).Bone marrow produces abnormal white blood cells(WBCs).These abnormal cells should die after some time but in reality they do not die and they become numerous in count. The normal white blood cells are interrupted by those abnormal white blood cells in doing their normal work. And this type of situation is named as disease like „Leukemia". Leukemia can be classified into Chronic and Acute leukemia.

1.1. Artificial Intelligence:

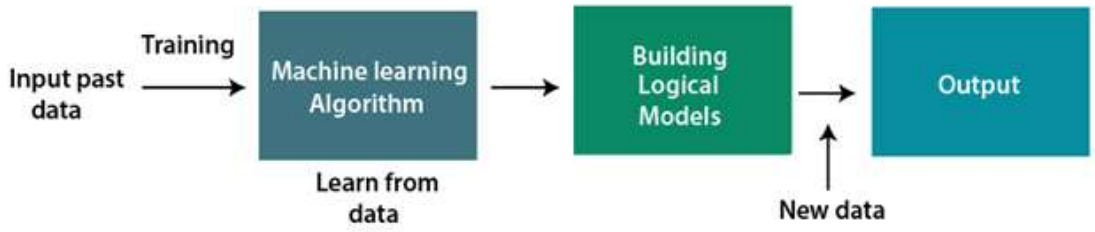
Artificial intelligence (AI) is the ability of a computer program or a machine to think and learn. It is also a field of study which tries to make computers "smart". As machines become increasingly capable, mental facilities once thought to require intelligence are removed from the definition. AI is an area of computer sciences that emphasizes the creation of intelligent machines that work and reacts like humans. Some of the activities computers with artificial intelligence are designed for include: Face recognition, Learning, Planning, Decision making etc., Artificial intelligence is the use of computer science programming to imitate human thought and action by analysing data and surroundings, solving or anticipating problems and learning or self-teaching to adapt to a variety of tasks.

1.2. Machine Learning

Machine learning is a growing technology which enables computers to learn automatically from past data. Machine learning uses various algorithms for building mathematical models and making predictions using historical data or information. Currently, it is being used for various tasks such as image recognition, speech recognition, email filtering, Facebook auto-tagging, recommender system, and many more. Machine Learning is said as a subset of artificial intelligence that is mainly concerned with the development of algorithms which allow a computer to learn from the data and past experiences on their own. The term machine learning was first introduced by Arthur Samuel in 1959. We can define it in a summarized way as: "Machine learning enables a machine to automatically learn from data, improve performance from experiences, and predict things without being explicitly programmed".

A Machine Learning system learns from historical data, builds the prediction models, and whenever it receives new data, predicts the output for it. The accuracy of predicted output depends upon the amount of data, as the huge amount of data helps to build a better model which predicts the output more accurately.

Suppose we have a complex problem, where we need to perform some predictions, so instead of writing a code for it, we just need to feed the data to generic algorithms, and with the help of these algorithms, machine builds the logic as per the data and predict the output. Machine learning has changed our way of thinking about the problem. The below block diagram explains the working of Machine Learning algorithm:



1.2.1. Features of Machine Learning:

- Machine learning uses data to detect various patterns in a given dataset.
- It can learn from past data and improve automatically.
- It is a data-driven technology.
- Machine learning is much similar to data mining as it also deals with the huge amount of the data.

1.2.2. Classification of Machine Learning

At a broad level, machine learning can be classified into three types:

1. Supervised learning
2. Unsupervised learning
3. Reinforcement learning

Supervised Learning

Supervised learning is a type of machine learning method in which we provide sample labeled data to the machine learning system in order to train it, and on that basis, it predicts the output.

The system creates a model using labeled data to understand the datasets and learn about each data, once the training and processing are done then we test the model by providing a sample data to check whether it is predicting the exact output or not.

The goal of supervised learning is to map input data with the output data. The supervised learning is based on supervision, and it is the same as when a student learns things in the supervision of the teacher. The example of supervised learning is **spam filtering**.

Supervised learning can be grouped further in two categories of algorithms:

- **Classification**
- **Regression**

Unsupervised Learning

Unsupervised learning is a learning method in which a machine learns without any supervision. The training is provided to the machine with the set of data that has not been labeled, classified, or categorized, and the algorithm needs to act on that data without any supervision. The goal of unsupervised learning is to restructure the input data into new features or a group of objects with similar patterns.

In unsupervised learning, we don't have a predetermined result. The machine tries to find useful insights from the huge amount of data.

It can be further classifieds into two categories of algorithms:

- **Clustering**
- **Association**

1.3. WBC:

A WBC count is a blood test to measure the number of white blood cells (WBCs) in the blood. WBCs are also called leukocytes. They help fight infections. White blood cells (WBC) are a heterogeneous group of nucleated cells that can be found in circulation for at least a period of their life. Their normal concentration in blood varies between 4000 and 10,000 per microliter. They play a most important role in phagocytosis and immunity and therefore in defense against infection. The simplest test is the WBC count and differential. White cells can be counted manually in specially designed chambers (Neubauer) or with automated counters. The latter are widely used, offering the advantage of higher accuracy and speed over manual techniques. To determine the differential, a drop of blood is thinly spread over a glass slide, air dried, and stained with a Romanofsky stain, most commonly the Wright or May-Grunewald-Giemsa technique. Two hundred cells are then counted and classified. Machines have been developed to perform automated differential counts, but they are still inferior to manual techniques as far as reliability and ability to discover morphologic abnormalities. The absolute number of each type of WBC, often more informative than its proportion, can be calculated if the differential and the total number of leukocytes per volume unit are known. Many of the conditions affecting the WBC can be diagnosed from studying the peripheral smear, but it may be necessary to evaluate the bone marrow for a better investigation. Bone marrow can be aspirated from the posterior iliac crest or the sternum. A core biopsy can be obtained percutaneously from the iliac crest. Biopsies allow assessment of the architecture of the marrow. Touch preparations can be made at the time the core bone marrow is obtained. Clumps of metastatic epithelial cancer cells can be recognized easily with this technique.

1.3.1. WBC is low

When you have a low white blood cell count, your immune system isn't working as well as it should. Doctors call this immune compromised. If you're immune compromised, you have a higher risk of getting an infection. White blood cells are produced by your bone marrow to help your body fight infection.

1.3.2. WBC is High

Produced in your bone marrow, they defend your body against infections and disease. But, when there are too many white blood cells, it usually means you have infection or inflammation in your body. Less commonly, a high white blood cell count could indicate certain blood cancers or bone marrow disorders.

II. LITERATURE REVIEW

[1] Title: Leukocyte segmentation and classification in blood-smear images

Author Name: H. Ramosernced Computer Vision GmbH-ACV,

Description: The detection and classification of leukocytes in blood smear images is a routine task in medical diagnosis. In this paper we present a fully automated approach to leukocyte segmentation that is robust with respect to cell appearance and image quality. A set of features is used to describe cytoplasm and nucleus properties. Pairwise SVM classification is used to discriminate between different cell types. Evaluation on a set of 1166 images (13 classes) resulted in 95% correct segmentations and 75% to 99% correct classification (with reject option)

[2] Title: A fast and efficient segmentation scheme for cell microscopic image

Author Name: G. Lebrun

Description: Microscopic cellular image segmentation schemes must be efficient for reliable analysis and fast to process huge quantity of images. Recent studies have focused on improving segmentation quality. Several segmentation schemes have good quality, but processing time is too expensive to deal with a great number of images per day. For segmentation schemes based on pixel classification, the classifier design is crucial since it is the one which requires most of the processing time necessary to segment an image. The main contribution of this work is focused on how to reduce the complexity of decision functions produced by support vector machines (SVM) while preserving recognition rate. Vector quantization is used in order to reduce the inherent redundancy present in huge pixel databases (i.e. images with expert pixel segmentation). Hybrid color space design is also used in order to improve data set size reduction rate and recognition rate. A new decision function quality criterion is defined to select good trade-off between recognition rate and processing time of pixel decision function. The first results of this study show that fast and efficient pixel classification with SVM is possible. Moreover

posterior class pixel probability estimation is easy to compute with Platt method. Then a new segmentation scheme using probabilistic pixel classification has been developed. This one has several free parameters, and an automatic selection must dealt with, but criteria for evaluate segmentation quality are not well adapted for cell segmentation, especially when comparison with expert pixel segmentation must be achieved. Another important contribution in this paper is the definition of a new quality criterion for evaluation of cell segmentation. The results presented here show that the selection of free parameters of the segmentation scheme by optimisation of the new quality cell segmentation criterion produces efficient cell segmentation.

[3] Title: An automated differential blood count system

Author Name: G Ongun

Description: While the early diagnosis of hematopoietic system disorders is very important in hematology, it is a highly complex and time consuming task. The early diagnosis requires a lot of patients to be followed-up by experts which, in general is unfeasible because of the required number of experts. The differential blood counter (DBC) system that we have developed is an attempt to automate the task performed manually by experts in routine. In our system, the cells are segmented using active contour models (snakes and balloons), which are initialized using morphological operators. Shape based and texture based features are utilized for the classification task. Different classifiers such as k-nearest neighbors, learning vector quantization, multi-layer perceptron and support vector machine are employed.

[4] Title: Leukocyte segmentation and classification in blood-smear images

Author Name: H. Ramoser

Description: The detection and classification of leukocytes in blood smear images is a routine task in medical diagnosis. In this paper we present a fully automated approach to leukocyte segmentation that is robust with respect to cell appearance and image quality. A set of features is used to describe cytoplasm and nucleus properties. Pairwise SVM classification is used to discriminate between different cell types. Evaluation on a set of 1166 images (13 classes) resulted in 95% correct segmentations and 75% to 99% correct classification (with reject option)

[5] Title: Routine bone marrow examination in the management of acute lymphoblastic leukaemia of childhood

Author Name: M K Palmer

Description: Eighty-four children with acute lymphoblastic leukaemia (ALL) who had relapsed in bone marrow were studied to assess whether treatment would be more successful if relapse was detected before the disease became clinically evident. Patients whose relapse was detected by routine bone marrow examination before the disease became apparent were compared with those whose relapse was suspected from clinical examination or peripheral blood findings. In the former there was a lower percentage of blast cells in the marrow (p less than 0.02) and the patients suffered less from complications of the disease, but there was no difference in the incidence or duration of second remissions between the two groups.

III. PROBLEM DEFINITION

- In our Existing system to detect the Leukemia based on microscopic images.
- The early and fast identification of Leukemia greatly aids in providing the appropriate treatment.
- Initial segmentation is done using Statistical parameters such as mean, standard deviation which segregates white blood cells from other blood components i.e. erythrocytes and platelets.
- Geometrical features such as area, perimeter of the white blood cell nucleusis investigated for diagnostic prediction of Leukemia.
- It successfully applied to a large number of images, showing promising results for varying image quality. Different image processing algorithms such as Image Enhancement, Thresholding, Mathematical morphology and Labelling are implemented using LabVIEW.

Drawbacks

- WBC Count not included.
- Inefficiency.
- Less Performance.

IV. DEVELOPMENT PROCESS

3.1. Requirement Analysis and Specifications

The requirement engineering process consists of feasibility study, requirements elicitation and analysis, requirements specification, requirements validation and requirements management. Requirements elicitation and analysis is an iterative process that can be represented as a spiral of activities, namely requirements discovery, requirements classification and organisation, requirement negotiation and requirements documentation.

3.1.1. Input Requirement and Output Requirements

Input Design

The input design is the link between the information system and the user. It comprises the developing specification and procedures for data preparation and those steps are necessary to put transaction data in to a usable form for processing can be achieved by inspecting the computer to read data from a written or printed document or it can occur by having people keying the data directly into the system. The design of input focuses on controlling the amount of input required, controlling the errors, avoiding delay, avoiding extra steps and keeping the process simple. The input is designed in such a way so that it provides security and ease of use with retaining the privacy. Input Design considered the following things:

- What data should be given as input?
- How the data should be arranged or coded?
- The dialog to guide the operating personnel in providing input.
- Methods for preparing input validations and steps to follow when error occur.

Objectives

1.Input Design is the process of converting a user-oriented description of the input into a computer-based system. This design is important to avoid errors in the data input process and show the correct direction to the management for getting correct information from the computerized system.

2.It is achieved by creating user-friendly screens for the data entry to handle large volume of data. The goal of designing input is to make data entry easier and to be free from errors. The data entry screen is designed in such a way that all the data manipulates can be performed. It also provides record viewing facilities.

3.When the data is entered it will check for its validity. Data can be entered with the help of screens. Appropriate messages are provided as when needed so that the user will not be in maize of instant. Thus, the objective of input design is to create an input layout that is easy to follow

Output Design

A quality output is one, which meets the requirements of the end user and presents the information clearly. In any system results of processing are communicated to the users and to other system through outputs. In output design it is determined how the information is to be displaced for immediate need and also the hard copy output. It is the most important and direct source information to the user. Efficient and intelligent output design improves the system's relationship to help user decision-making.

1. Designing computer output should proceed in an organized, well thought out manner; the right output must be developed while ensuring that each output element is designed so that people will find the system can use easily and effectively. When analysis design computer output, they should Identify the specific output that is needed to meet the requirements.

2.Select methods for presenting information.

3.Create document, report, or other formats that contain information produced by the system.

The output form of an information system should accomplish one or more of the following objectives.

- Convey information about past activities, current status or projections of the
- Future.
- Signal important events, opportunities, problems, or warnings.
- Trigger an action.
- Confirm an action.

V. PROPOSED METHOD

- In Our Proposed method, the paper implemented the main point of automatic counting of WBCs to detect leukemia automatically from the image which is taken from the microscope.
- The manual counting process to detect leukemia is very much time taking and gives inaccurate result also. So to save lives this automated proposed method is very important.
- Here, the Image Processing Techniques will be implemented based on Machine Learning. Some Steps will be implemented to show the results.
- Datasets collection, Preprocessing, Segmentation, WBC Count, Feature Extraction and Classifications are the steps will be demonstrated. Experimental results shows the better performance when compared to existing method.
- Finally to compare all the algorithm to show the result in which algorithm is best

4.1. Advantages

- Better Performance.
- Accurate Results.
- WBC Count can be measured.
- Time Consumption.

4.2. Algorithm

Ensemble:

Ensemble methods is a machine learning technique that combines several base models in order to produce one optimal predictive model . To better understand this definition lets take a step back into ultimate goal of machine learning and model building.

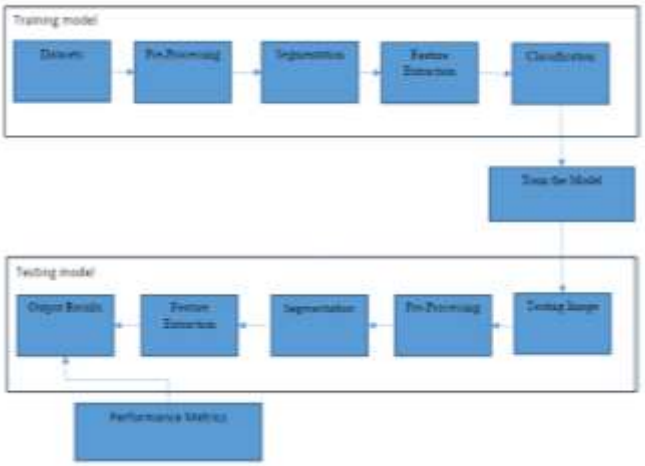
SVM:

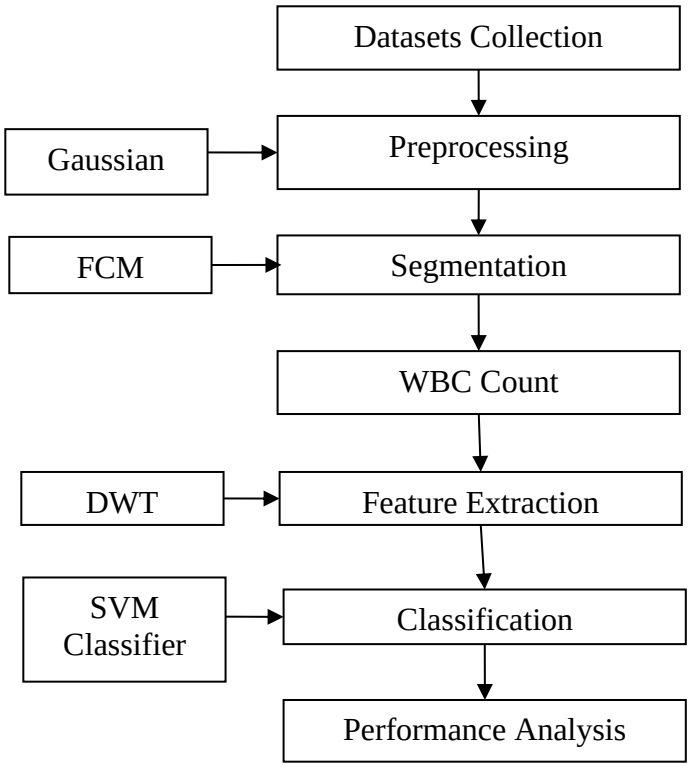
Support vector machines (SVMs) are a set of supervised learning methods used for classification, regression and outliers detection. The advantages of support vector machines are: Effective in high dimensional spaces. Still effective in cases where number of dimensions is greater than the number of samples.

FNN:

Feed forward networks consist of a series of layers. The first layer has a connection from the network input. Each subsequent layer has a connection from the previous layer. The final layer produces the network’s output. feed forward networks consist of a series of layers. The first layer has a connection from the network input. Each subsequent layer has a connection from the previous layer. The final layer produces the network's output. You can use feed forward networks for any kind of input to output mapping.

VI. SYSTEM ARCHITECTURE





4.3. Modules Used

- In this Project has five modules:
- 1. Datasets Collection.
- 2. Pre-processing.
- 3. Segmentation.
- 4. Feature Extraction.
- 5. Classification.

1. Datasets Collection:

A data set (or dataset) is a collection of data. In the case of tabular data, a data set corresponds to one or more database tables, where every column of a table represents a particular variable, and each row corresponds to a given record of the data set in question.

2. **Preprocessing:**

Pre-processing routines prepare the data for analysis. Before we start the actual processing, the data has to be pre-processed to remove the detector effects. Preprocessing is the most important aspect of data processing. Hence, data filtering, data ordering, data editing and noise modeling play an important role in any data preprocessing.

3. Segmentation:

Image segmentation is the process of partitioning an image into parts or regions. This division into parts is often based on the characteristics of the pixels in the image. For example, one way to find regions in an image is to look for abrupt discontinuities in pixel values, which typically indicate edges.

4. Feature Extraction:

Feature extraction refers to the process of transforming raw data into numerical features that can be processed while preserving the information in the original data set. It yields better results than applying machine learning directly to the raw data.

5. Classification:

Classification is a type of supervised machine learning in which an algorithm "learns" to classify new observations from examples of labeled data. Here we can use Ensemble, SVM, FNN algorithm to classify the data based on user selecting image.

VII. FUTURE ENHANCEMENT

In our future work, we can implement the concept of automatic leukemia detection disease using image processing in deep learning techniques in future.

VIII. CONCLUSION

In this paper the main point of concern is automatic counting of WBCs to detect leukemia automatically from the image which is taken from the microscope. The manual counting process to detect leukemia is very much time taking and gives inaccurate result also. So to save lives this automated proposed method is very important. Actually, in two ways we can detect leukemia. one is by automatic counting and on the other way SVM and Ensemble and FNN classifier will tell automatically whether the image has leukemia effected cells or not. For the first one after segmentation counting algorithm is used to count automatically. And for the second one some features are extracted to form a feature matrix to compare with the standard values. We have achieved an accuracy of almost high using this proposed method of counting.

REFERENCES

[1] H. Ramoser, V. Laurain, H. Bischof and R. Ecker, "Leukocyte segmentation and classification in blood-smear images," in IEEE Engineering in Medicine and Biology, Shanghai, 2005.

[2] G. Lebrun, C. Charrier, "A fast and efficient segmentation scheme for cell microscopic image," in Cellular and Molecular Biology TM 53, N°2, 51-61, 2007 .

[3] Marco Antonio Garcia de Carvalho ,Tiago William Pinto, Roberto Marcondes C´esar J´unior," Image Segmentation Using Watershed and Normalized Cut".

[4] G. Ongun, U. Halici, K. Leblebicioglu and V. Atala, "An automated differential blood count system," in Int. Conf. of the IEEE Engineering in Medicine and Biology Society, 2001, Istanbul.

[5] Mohapatra, Saurav, Dipti Patra, Sudhakar Kumar, and Siddhartha Satpathi. "Kernel induced rough c-means clustering for lymphocyte image segmentation." In Intelligent Human Computer Interaction (IHCI), 2012 4th International Conference on, pp. 1-6. IEEE, 2012 .

[6] Raje, Chaitali, and Jyoti Rangole. "Detection of Leukemia in microscopic images using image processing." In Communications and Signal Processing (ICCSP), 2014 International Conference on, pp. 255-259. IEEE, 2014.

[7] Berge, Heidi, Dale Taylor, Sriram Krishnan, and Tania S. Douglas. "Improved red blood cell counting in thin blood smears." In Biomedical Imaging: From Nano to Macro, 2011 IEEE International Symposium on, pp. 204-207.

[8] Abdul Nasir, A. S., M. Y. Mashor, and H. Rosline. "Unsupervised colour segmentation of white blood cell for acute leukaemia images." In Imaging Systems and Techniques (IST), 2011 IEEE International Conference on, pp. 142-145. IEEE, 2011