

Acid Fast Stain Sputum Smear Images Data Set for Diagnosis of Tuberculosis

Mukti Jadhav
Institute of Management
Studies & Information
Technology Aurangabad,

Prachi P. Janrao
Thakur Engineering College,

K.V. Kale
Department of Computer
Science & IT, Dr.
BabasahebAmbedkarMarathw
ada University, Aurangabad.

ABSTRACT

Many of the computer science researchers now work for design automatic system for fast & accurate recognition of tuberculosis(TB), to speed up the treatment process of patient to save patient life. Any type of research is not possible without data. It is critical for computer researcher to prepare Ziehl-Neelsen stain sputum smear slide & to setup image acquisition technique on simple light microscope. We are preparing acid fast stain sputum smear images data set using light microscope with digital camera at the eye piece of microscope. The dataset contain 400 total images (300 images are positive for Tb and 100 negative for Tb). These captured image dataset is available for many researchers. One of the application is applied on 50 of these images

KEYWORDS

Ziehl-Neelsen stain, image segmentation, Acid fast stain.

1. INTRODUCTION

In the world tuberculosis is leading cause of death. To control TB early and accurate diagnosis is require. The fast recognition of TB improves treatment outcomes, reduces transmission of TB and allows for greater control of drug resistance and multi drug resistant TB. Sputum smear examination through smear microscopy is universally accepted for laboratory diagnosis of tuberculosis, as well as the principal examination to monitor tuberculosis treatment. Sputum examination by microscopy is relatively quick, easy, and inexpensive [1]. In developing countries diagnosis and treatment of TB is using smear microscopy with the Ziehl-Neelsen (ZN Staining) technique, due to its simplicity, rapidity, reproducibility, low cost, and effectiveness in detecting infectious cases [2]. Mycobacterium tuberculosis excreted in sputum is visible in clusters or individually in stained sputum smears under a microscope is red in color over blue background with ZN stain (Acid Fast Stain). Purpose of ZN Stain

It used in the demonstration of acid-fast bacteria belonging to the genus 'mycobacterium', which include the causative agent for tuberculosis [3].

"Permission to make digital or hard copies of part or all of this work for personal or classroom use is granted without fee provided that copies are not made or distributed for profit or commercial advantage and that copies bear this notice and the full citation on the first page. Copyrights for components of this work owned by others than IJCA must be honored. Abstracting with credit is permitted. To copy otherwise, to republish, to post on servers or to redistribute to lists, needs an acknowledgement

A. Acid Fast Staining Principle

Mycobacterial cell walls contain a waxy substance composed of mycolic acids. These are β -hydroxy carboxylic acids with chain lengths of up to 90 carbon atoms. The property of acid fastness is related to the carbon chain length of the mycolic acid found in any particular species [4].

Basic fuchsin binds to negatively charged groups in bacteria. The mycolic acid (and other cell wall lipids) present a barrier to dye entry as well as elusion (washing out with solvent) and this is partly overcome by adding a lipophilic agent to a concentrated aqueous solution of basic fuchsin and partly by heating.

B. Reagent Formulae

1. Staining solution

Stock Solution A (stable for 6 months)

L.O.C. High Suds (Amway) 0.6 ml

Distilled water 100 ml

Stock Solution B

Basic fuchsin 1 g

Absolute ethyl alcohol 10 ml

The two solutions can be kept as stock solution and mixed before use

2. Working Solution (stable for 1 month)

Mix 50ml of A with 5 ml of B.

3. 3% hydrochloric acid in 95% ethyl alcohol

Absolute ethyl alcohol 95ml

Distilled water 2 ml

Concentrated hydrochloric acid 3 ml

Make up the alcohol solution then add the concentrated acid.

Use extreme care when handling concentrated acid.

4. 0.25% methylene blue in 1% acetic acid

Methylene blue 0.25 g

Distilled water 99 ml

Acetic acid 1 ml

2. METHODS

A. Preparation of slide & sputum smear -

Select good glass slide, clean and dry, Heat end gently, Mark number outer third portion. Then select mucoid portion of patient sputum spread evenly on middle 1/3 slide, allow smear to air dry, gently pass over flame thrice, flame wire loop and burn used stick.

B. Preparation of slide with ZN stain -

Place slide on Rack, Smear side up, cover smear with filter paper. Heat fix sputum on glass slide. Flood the slide with carbolfuchsin stain. Heat the slide gently until it steams (5min). Pour off the carbolfuchsin. Wash slide thoroughly with

water. Decolorized with acid-alcohol (5min). Wash slide thoroughly with water. Flood slide with methylene blue counter stain for 1 min. Wash with water. Blot excess water and dry in hand over bunsen flame. The Slide is ready to observe under microscope.

C. Examination of standard smear for ABF

Place 1 drop of cedar wood oil/Liquid paraffin. Mount slide on Microscope. Adjust mirror to light source. Raise the condenser, Adjust light reflection into condenser. Lower oil immersion objective(100x), view sideways.Allow objective to touch oil-drop, use coarse adjustment. Bring field into focus use fine adjustment[5].Examine as per grid pattern shown in following figure.

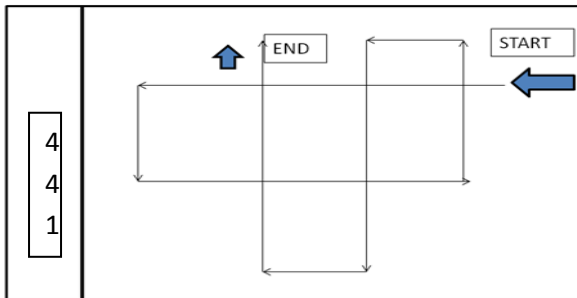


Figure.1 Grid pattern for sputum examination.
Like this way examine 100 field record results.

4. IMAGE ACQUISITION

Images are capture with the help of lightmicroscope, digital camera and computer system. A camera is on the eyepiece position of microscope.It is directly connected to the computer system. A camera having inbuilt software for capturing and storing images.The microscope magnification is computed by multiplying the objective’s scale number by the eye piece magnification. Objective is 100x and 2 mega pixel, 2/3 inch digital camera is attached to the eyepiece of the microscope MP. The illumination is effected with the aid of the 15 W main connecting lamp. These images can directly display on the monitor screen if it is clear then captured and save in to bitmap (*.bmp) format on the desired location of memory.

DIGITAL CAMERATABLE 1 : DCE-2 SPECIFICATION

Image Sensor	1/4"CMOS
CMOS Size	10.2mm×10.2mm
Valid Pixel (1420)	640×480(0.3M Pixel)
Pixel Size	5.04 um×5.04um
Frame Rate	30frame/s(VGA)
Image Format (1420)	640×480
	352×288
	320×240
	176×144
	160×120
Image Output	USB1.1 and USB2.0 ,Plug and Unplug Freely.
Digital Output	10-bit
Lowest Illumination	1.2lux

SNR	>44dB
Dynamic Range	60dB
Camera Image Control	Saturation、 Contrast、 Sharpness and so on
White Balance	Auto、 Manual
Exposure	Auto、 Manual (1~500ms)
Working Temperature	--20℃ ~+20℃
Dimensions	26.80mm×16.00mm
Connecting Mode	Insert it into Eyepiece Tube of Microscope Directly or Using Standard C Mount
Image Output	Plug and Unplug Freely, Automatic Power Controlling

With high-resolution 640×480, avoid mosaic phenomenon and make image fluen. Provide auto brightness and white balance control function and some high-level image control function such as color saturation, contrast fringe swell and gamma numerical value.

Adopt high performance USB2.0 chip. Convenient to use and just insert into eyepiece tube or top tube of trinocular head. Easy to install and operate and it can be plug unplug freely.

Sensors: The heart of every digital camera is the Sensor, usually either a CCD or a CMOS type. All Sensors are analog devices, converting photons into electrical signals. The process by which the analog information is changed to digital is called Analog to Digital (A/D) Conversion. This can occur within the Sensor itself, or within the camera, or within the PC, as was the case with older video camera / frame grabber technology. The Sensor size is another consideration. The larger the Sensor size (actually the larger the individual pixel size), the more light sensitive the Sensor should be, so larger is better, at least in theory

Resolution: The Sensor Resolution is the total number of picture elements (Pixels) forming the image

5. SYSTEM REQUIREMENTS

- Operating System Windows 98/me/2000, Windows XP or Windows Vista.
- DirectX 9.0 or higher (for Windows 98/me or 2000)
- Video adapter supports 24bit color or more
- CPU with 1.5GHz or more.
- System Memory 128MB or more, Display Memory 64MB or more.
- USB2.0 interface.
- Hard Disk Space 1GB for installation plus additional space for captured images

Since video processing is hardware intensive, a faster computer with a fast hard disk drive and extra memory will yield better results.



Figure.2. Setup for image acquisition.

6. IMAGE SPECIFICATION

Type of file Bitmap image[6], Width 640 pixels, Height 480 pixels (640 x 480) Horizontal resolution 96 dpi, Vertical resolution 96 dpi Bit depth 24, frame count 1 Color True-color, file size 900 KB, Size on disk 904 KB uncompressed size 900 KB, Compressed ratio 1. What's the difference in regular camera and DCE-2 camera ?

Dedicated microscope digital cameras typically cost more for less resolution, but they have the advantage of being directly controlled by a computer. They typically provide a better preview image, often with focusing aids, and the ability to acquire the image directly. There are some drawbacks while capturing these images by new users & hand camera.

The non biological researcher can't operate microscope finely, they can't differentiate stain residue and tuberculosis bacterial cells. While capturing images from microscope with hand camera due to hand vibration image may become blur.

7. APPLICATION

- These digital images are useful for researchers work on lung disease, clinical pathologists.
- Many computer science researchers related with medical image processing, pattern recognition, Neural Network etc. yielding much more flexibility and time savings with these images.
- Images can be used for TB screening aims to speed up the screening process, to improve its sensitivity and to reduce its reliance on technicians.
- Comparison between two different images is possible.

The image data set contains total 400 images with 300 TB positive images and 100 TB negative images. The data set is totally free for many researchers. It can be accessed by requesting to the author by mailing, some of the sample images are available on the given site <http://imsitaurangabad.com/images/TBsputumimages>

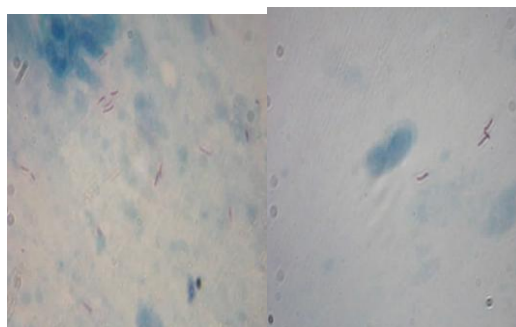


Figure.3 Sputum smear image Figure.4 Sputum smear image

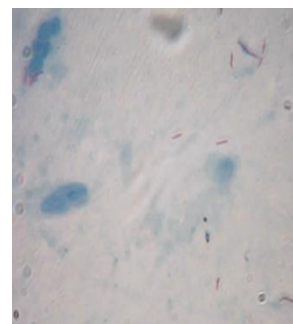


Figure.5 Sputum smear image

ZN STAIN RESULTS

Acid fast bacilli - Red

Nuclei - Blue

Other tissue constituents - Blue

8. COMPUTERIZE DIAGNOSIS OF TB

The manual screening method is time consuming, tedious & sometime it may be confusing with some non-tuberculosis bacilli or some rod-shaped stain residue. Due to this we get faulty results. For this reason, the need of automation is required. The captured images of sputum smear need to go through various processes for proper diagnosis of TB. In this proposed work, 50 images of sputum smear are analyzed. These images are enhanced for better color contrast and noise reduction, then segmented on the basis of color. After segmentation, 219 objects are extracted. These objects are recognized for M.TB bacilli cells, stain residues, non-TB cells for diagnosis of TB. The recognized objects from sputum smear images are counted to find out the level of TB as per WHO provided chart means 1+, 2+, 3+, scanty or negative. This becomes helpful to lab technicians for fast and accurate TB diagnosis. Similarly, these results can be stored for a particular patient for future and further diagnosis, like MDR-TB or XDR-TB.

Smoothing: Suppress as much noise as possible, without destroying the true image.

Enhancement: Apply filter like contrast stretching to enhance the quality of the sputum smear images for clear diagnosis of M.TB bacilli in the image. The enhanced images are visually enhanced than the original image and clearly differentiate reddish pink color M.TB bacilli on blue background sputum smear.

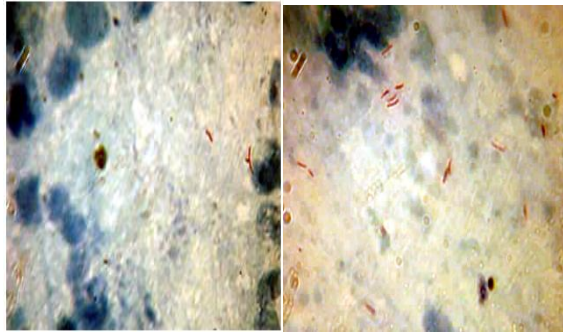


Figure.6 Sputum smear image Figure.7 Sputum smear image



Figure.8 Sputum smear image

Quality Measures :- A good objective quality measure should well reflect the distortion on the image due to blurring, noise, compression, sensor inadequacy[7]. The quality of image is measured by statistical mean, standard deviation & histogram

9. Table 2: Quality measures for image before & after image enhancement.

Name of image	Mean		Standard Deviation	
	Original	Contrast Stretch	Original	Contrast Stretch
1	161.526	167.8935	8.6967	46.4862
2	163.674	169.5572	8.9067	48.116
3	159.935	170.3907	7.2944	52.5222

Table shows the statistical quality measures for digital images of ZN-stain sputum smear images after enhancement of images using Contrast Stretching (Normalization) identification of M.TB Bacillicells.

Histogram: Histogram represents statistical analysis of an image. The histogram is a bar graph. Each vertical bar in the graph indicates the number of pixels of that value. The sum of all vertical bars is equal to the total number of pixels in the image[8]It indicates the distribution of image data values using these values we can determine the following things:-

- The best contrast image.
- The overall brightness or darkness of an image. How to apply a function, such as thresholding to an image.

We can judge the effectiveness of contrast enhancement, applied to an image

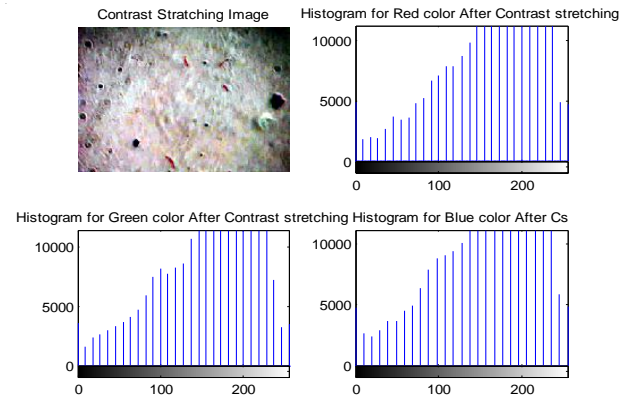


Figure.9 Contrast stretching image Histogram for Red, Green and Blue color

Segmentation : Segmentation process subdivides an image into its constituent regions or objects. Image segmentation algorithms generally are based on one of the two basic properties of intensity values: discontinuity and similarity. Due to Acid Fast Staining color is the discriminate feature of M.TB bacilli, which is segmented by using thresholding method from the background sputum smear.

The segmented image shows two values either 255 (white pixel) which indicate the area of M.TB pixel or 0 (black pixel) sputum area. We get final value which is either 255 (white pixel) or 0 (black pixel) i.e M.TB cells are white in color on black color background. The obtained image contains M.TB cells with some stained residues, non decolorized epithelial cells and non TB bacillus.

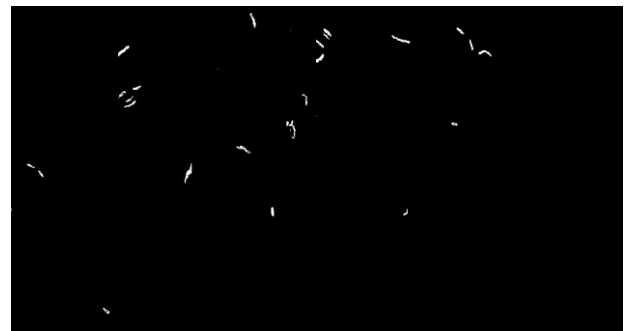


Figure.10&Figure.11 Sputum smear image after segmentation

Edge Detection

The proper recognition of M.TB bacilli from the sputum smear images the boundary detection technique is applied on the extracted object after segmentation. The extracted objects processed from morphological operation such as dilation. Then the shape based features of these objects are extracted.



Figure.12 Various shapes object are obtained after segmentation and boundary detection

Feature Extraction

For exact recognition of M.TB.bacilli we need to extract some features of the organisms. Shape is a fundamental property of an organism i.e. object, The shape descriptor moment invariants are one of the most popular and widely accepted & used contour-

based shape descriptors is a set of derived by Hu[9]. Another shape based features are eccentricity, perimeter, and area are extracted.

Moment invariant

The recognition of objects regardless of their position, size and orientation seven values are derived. These values computed by normalizing central moments through order three, that are invariant to object scale, position, and orientation.

Eccentricity

The eccentricity is the ratio of the distance between the foci of the ellipse and its major axis length [10].

$$e = \frac{\sin \alpha}{\sin \beta}$$

Where α is the angle between the plane and the horizontal and β is the angle between the cone and the horizontal. For ellipse

eccentricity “e” is calculated as
$$e = \sqrt{1 - \frac{b^2}{a^2}}$$

Where a is the length of its semi-major axis and b be the length of its semi-minor axis.

Perimeter

It is p-element vector containing the distance around the boundary of each contiguous region in the image. Perimeter of various shapes.

Area General formula :Area is a quantity that expresses the extent of a two-dimensional surface or shape in the plane. The general formula for the surface area of the graph of a continuously differentiable

functions $Z = f(x, y)$ where $(x, y) \in DC \mathbb{R}^2$ and Dis a region in the xy-plane with the smooth boundary:

$$A = \iint_D \sqrt{\left(\frac{\partial f}{\partial x}\right)^2 + \left(\frac{\partial f}{\partial y}\right)^2} + 1 dx dy$$

Table 3:Shape base features are calculated.

Perimeter	Eccentricity	Area
51.55635	0.91553019	60
81.15433	0.95344361	65
75.94113	0.9321375	66
54.42641	0.9468712	41
102.468	0.84938513	85
119.7817	0.93980161	50
53.79899	0.9493625	48
89.25483	0.93008035	76
38.97056	0.51586125	34

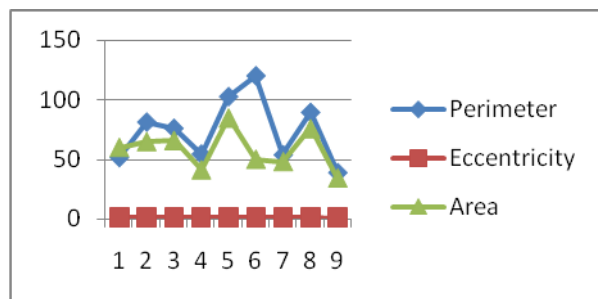


Figure.13Graph shows Shapebase features of objects.

Table 4: Moment invariants(M1 – M7) are used as a feature

M1	M2	M3	M4	M5	M6	M7
0.280608	1.206832	3.124399	3.74719	7.426976	4.588074	10.51907
0.593288	0.824618	0.989948	1.213587	4.458548	2.302337	5.5447
0.138538	0.251597	2.84831	2.36403	6.066325	3.398834	9.767206
0.578826	0.960896	1.676451	1.043145	3.868302	1.602652	6.802666
0.051034	1.242361	1.954156	5.137873	9.035852	8.185343	9.210848

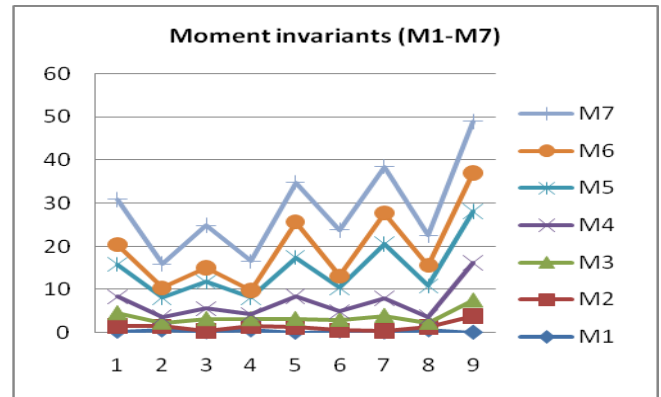


Figure.14 Graph shows Moment invariants(M1 – M7) are used as a feature

Pattern Classification: The extracted objects are classified using classifiers. M.TB bacilli and NonTB bacilli cells are recognized on the basis of features.

Support Vector Machines (SVM)

Support Vector Machines (SVM) is a technique for data classification. It is method which analyzes data and recognize for patterns, used for classification and regression. SVM is linear classifier often giving superior classification results than any other classifier. A classification task usually involves separating data into training and testing sets. In the training set contains one target value i.e. the class labels and several attributes i.e. the features variables. The goal of SVM is to produce a model (based on the training data) which predicts the target values of the test data given only the test data attributes [11]. SVM classifier design is the notion of the margin [12].

7k-Nearest Neighbor (KNN)

In pattern recognition, the **k-nearest neighbor algorithm** (k-NN) is a method for classifying objects based on closest training examples in the feature space. k-NN is a type of instance-based learning, or lazy learning where the function is only approximated locally and all computation is deferred until classification. The k-nearest neighbor algorithm is amongst the simplest of all machine learning algorithms: an object is classified by a majority vote of its neighbors, with the object being assigned to the class most common amongst its k nearest neighbors (k is a positive integer, typically small). If k = 1, then the object is simply assigned to the class of its nearest neighbor [13].

Nearest neighbor rules in effect compute the decision boundary in an implicit manner. It is also possible to compute the decision boundary itself explicitly, and to do so in an efficient manner so that the computational complexity is a function of the boundary complexity [14].

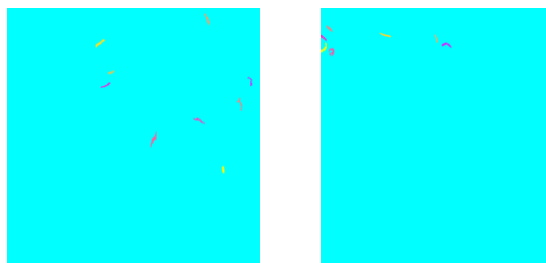
The nearest neighbor algorithm has some strong consistency results. As the amount of data approaches infinity, the algorithm is guaranteed to yield an error rate no worse than twice the Bayes error rate [15]. k-nearest neighbor is guaranteed to approach the Bayes error rate, for some value of k. Various

improvements to k -nearest neighbor methods are possible by using proximity graphs[16].

Table 5:Result obtained after applying classifier for recognition of M.TB bacilli.

Classifier	Training data	Testing data	Features	Classes	Accuracy	Specificity
SVM	219	169	10	2	85%	98.81%
KNN	219	160	10	2	88%	94.00%

The recognized objects are counted to assign grade mean level to the TB patient as per WHO chart. These steps enable to learn about the characteristics of the M.TB bacilli cells and easily compute statistics for all the bacilli in the image. Identify objects (recognized M.TB cells) in the Image, examine each object, view all the objects, compute area of each object, compute area-based Statistics and counting cells.



ACKNOWLEDGEMENTS

Special thanks to Dr. Deshpande S. District TB Officer, Mr. Jadhav B.S and Mr. Tarte Tuberculosis Lab Technician, TB Hospital Aurangabad.

REFERENCES

- [1] "Supporting Laboratory Services", (DHHS/CDC) funded by the U.S. Agency for International Development. Grant Agreement 118-G-00-99-00112 (WHO) and PASA 118-P-00-98-00165, WHO/CDS/TB/2002.310
- [2] J. A. C. Luna, "A Tuberculosis Guide for Specialist Physicians", International Union against Tuberculosis and Lung Disease, 2004.
- [3] "Surgical Pathology – Histology Staining Manual Microorganisms", Downloaded from WebPath: Internet Pathology Laboratory

<http://www-medlib.med.utah.edu/WebPath/webpath.html>

- [4] Lyon, H (1991), "Theory and strategy in histochemistry", Springer - Verlag
- [5] "Sputum Microscopy Instruction to Laboratory Technicians", Tuberculosis research center, Chetput. Printed at Diocesan press Madras-60007-C624
- [6] Matlab Help Files "Analyzing images", Math Works Cambridge MA. 2009
- [7] Smail Avci bas, Bulent Sankur, Khalid Sayood "Statistical evaluation of image quality measures", 206 / Journal of Electronic Imaging / April 2002 / Vol. 11(2)
- [8] Image Statistics Based on material from Digital Imaging: Theory and Applications, H. E. Burdick, McGraw-Hill, 1997
- [10] Rethabile Khutlang, Sriram Krishnan, Andrew Whitelaw, Tania S Douglas, "Detection of Tuberculosis in sputum smear images using two one-class classifiers", 978-1-4244-3932-4/09/\$25.00 ©2009 IEEE
- [9] Muharrem Mercimek, Kayhan Gulez and Tarik Velimurcu, "Real object recognition using moment invariants", Sadhana Vol. 30, Part 6, December 2005, pp. 765–775. © Printed in India
- [11] Chih-Wei Hsu, Chih-Chung Chang, and Chih-Jen Lin, "A Practical Guide to Support Vector Classification", Initial version: 2003 Last updated: April 15, 2010
- [12] Sergios Theodoridis Konstantinos, "An Introduction to Pattern Recognition A MATLAB Approach", British Library Cataloguing-in-Publication Data.
- [13] Bremner D, Demaine E, Erickson J, Iacono J, Langerman S, Morin P, Toussaint G, "Output-sensitive algorithms for computing nearest-neighbor decision boundaries". Discrete and Computational Geometry 33 (4): 593–604, 2005.
- [14] Cover TM, Hart PE (1967). "Nearest neighbor pattern classification". IEEE Transactions on Information Theory 13 (1): 21–27
- [15] Toussaint GT (April 2005). "Geometric proximity graphs for improving nearest neighbor methods in instance-based learning and data mining". International Journal of Computational Geometry and Applications 15 (2): 101–150.
- [16] Howard Anton (Feb. 2002), "Elementary Linear Algebra 5e" Publisher John Wiley & Sons Inc, ISBN 0-471-85223-6