Indian Journal of Engineering

ANALYSIS

ISSN 2319 – 7757 EISSN 2319 –7765

Hybrid PSO – Genetic Approach for Leukemia Segmentation

Publication History
Received: 04 August 2015
Accepted: 21 August 2015
Published: 1 September 2015 (Special issue)

Citation
Charles K, Hannah Inbarani H. Hybrid PSO – Genetic Approach for Leukemia Segmentation. *Indian Journal of Engineering*, 2015, 12(30), 60-66

Hybrid PSO – Genetic Approach for LeukemiaSegmentation

K.charles,
Research Scholar,
Department of Computer Science,
Periyar University, Salem-636011,
Tamilnadu, India.

Dr.H.Hannah Inbarani,
Assistant Professor,
Department of Computer Science,
Periyar University, Salem-636011,
Tamilnadu, India.

***Abstract*—** Microscopic analysis of stained blood slides is crucial method for identification of hematological disease. Human visual examinations of blood slides are slow and also it's depends upon specialist experience. Leukemia is a foremost reason for death around worldwide and affects both children's and adult's malignant neoplasm of the blood or bone marrow. To avoid this problem we use acute hymphoblastic leukemia(ALL) for save human life. We find its through Microscopic. This method is very useful to improve the all diagnostic accuracy by analyzing morphological and textual features from the blood image. Hybrid clustering based on PSO. The proposed dataset contain ALL-IDB2. It was accomplished through segmentation, feature extraction. To improve segmentation methodology using Genetic PSO algorithm.

Keywords: Acute lymphoblastic leukemia _ Morphological_ Genetic PSO.

1. INTRODUCTION

According to the American Cancer Society, cancer or malignant neoplasm is the world's leading cause of death followed by cardiovascular diseases. Cancer can be group of diseases characterized by: (i) uncontrolled cell division which prohibits programmed cell death and contributes to abnormal growth of tissues, (ii) ability to metastasize (spread), and (iii) eventually compromising the cellular function of the person, which successively may lead to death [1]. Cancer can affect any part of the body, although some cancers are more common or less common than others. According to the Centers for Disease Control and Prevention, 12.7 million people find out each year around the world that they have cancer and 7.6 million people die from cancer. And as per the joint study conducted by Centre for Global Health Research at St. Michael's Hospital, Toronto, and Indian national institutions in India, cancer alone accounted for 8 % of the 2.5 million total male deaths and 12 % of the 1.6 million total female deaths in the year 2010 [2].

Hematological malignancies are heterogeneous group of diseases which includes various forms of leukemia, lymphoma, and myeloma and are characterized by the malignant uncontrolled growth of hematopoietic cells [3]. The development of such malignancies results from an accumulation of genetic mutations in genes involved in regulating cell differentiation and proliferation, leading to aberrant control of these processes. It has been reported that approximately 75,000, 45,000, and 20,500 persons were diagnosed with lymphoma, leukemia, and myeloma, respectively, in 2011 in the USA alone [4]. In India, for the year 2010 approximately the total number of individuals suffering from blood cancer was estimated to be 104,239 [5]. And according to Indian Council of Medical Research (ICMR) by the year 2020, the total number of cancer cases of lymphoid and hematopoietic system is expected to go up to 77,190 for males and 55,384 for females. Even though leukemia starts in the bone marrow and lymphoma in the lymphatic system, both are considered as malignancies of the blood. They can affect people of all ages; however, leukemia is more common in children and young adults and people over the age of 60. The majority of leukemia deaths occur in low- and middle-income countries including India, where most of the patients are diagnosed in later stages. In India, leukemia is the most common childhood cancer with relative proportion varying between 25 and 40 % [6] and is the present subject of our study.

Definite genetic processes contribute toward malignant transformation of cells and their progeny forming a clone of leukemic cells [7]. Such neoplastic proliferations of hematopoietic cells are known as leukemia. Based on the severity of the disease, leukemia can be acute or chronic. Acute leukemia can be defined as neoplasms with more than 20 % of blasts in the peripheral blood/bone marrow and is a group of disorders which, if untreated, results in death in few weeks.

– Acute lymphoblastic leukemia (ALL)

Due to advancement in treatment modalities, it is always necessary to subclass the leukemia to assess the prognosis and for the suitable planning of the treatment. The

most widely used protocols for leukemia sub categorization are World Health Organization (WHO) classification and French, American, British (FAB) [8]. But, both fundamentally divide leukemia's as myeloid and lymphoid types, depending on the origin of the blast cell. Acute lymphoblastic leukemia (ALL) is the single most common pediatric malignancy accounting for one-fourths of all childhood cancers, thus considered as our current research focus. ALL affects both children and adults; however, primarily it is a childhood disease with peak prevalence between the age of 2 and 5 years. According to WHO, ALL subtypes are based on whether the precursor cell is a T or B lymphocyte, whereas FAB classification of ALL is based on morphology and histochemical staining and can be L_1, L_2, or L_3 subtypes.

Currently, microscopic examination of blood samples (peripheral blood/bone marrow) is a standard procedure for a confirmative screening and subtyping of ALL. However, regardless of advanced techniques like flow cytometer, immunophenotyping, and molecular probing, morphological evaluation of stained blood films still remains an economical procedure for the initial screening of ALL [9] across the globe. ALL diagnosis involves distinguishing a healthy lymphocyte from a malignant lymphocyte (lymphoblast) and can be difficult, even for an expert hematologist if the morphological features are not well developed or partially present. Nevertheless, there is always a chance of variability in human-reported diagnosis due to several factors, i.e., improper manual staining, operator fatigue, and inter-observer and intra-observer differences. Analysis of blood samples for hematological inferences is purely qualitative and is based on clinicopathological experience of the observer. As is the case at most regional cancer centers in India, visual diagnosis is often time-consuming and cumbersome as the number of cases per day is quite high across the country. Due to the prevalence of such uncertainty in manual screening of ALL, the conventional hematological evaluation needs to be strengthened using quantitative microscopy. Such automated procedures aim at avoiding painful biopsies and will facilitate early and precise diagnosis of leukemia. The representative blood microscopic images consisting of a lymphocyte (healthy) and a lymphoblast (malignant lymphocyte) are depicted in Fig. 1.

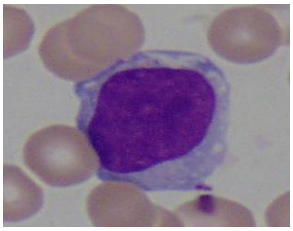

Fig. 1 Representative blood microscopic images containing a lymphocyte.

A brief overview of various standard classifiers is

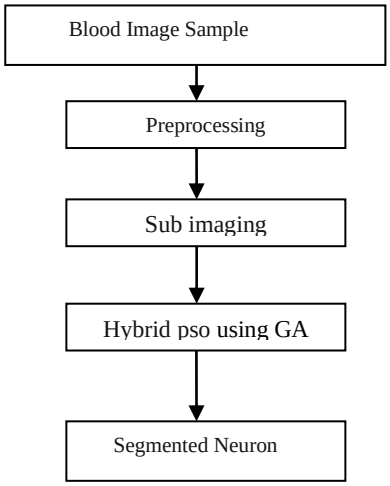

Fig. 2. The proposed framework of the WBC segmentation scheme.

2. METHODS

2.1 Image acquisition

Blood microscopic images of Leishman [11]-stained peripheral blood or bone marrow samples were optically grabbed by Zeiss Observer microscope (Carl Zeiss, Germany) under 100X oil-immersed setting and with an effective magnification of 1,000 at Ispat General Hospital, Rourkela, India. Each grabbed digital image is represented using three fundamental colors (red, green, and blue) and is stored in an array of size 1,024 X 1,024.

2.2 Preprocessing

The presence of noise and acquisition of blood microscopic images under uneven lighting conditions necessitates preprocessing. This is achieved using filtering and contrast enhancement.

2.3 Sub imaging

Peripheral blood smear images are relatively larger with more than one leukocyte per image Convert image from RGB Color space to L*a*b* .3d image into 2d image
Image segmentation of blood images is the foundation for all automated image-based hematological disease recognition systems including ALL. Image segmentation is performed in

L*a*b* (CIELAB) color space. This color space consists of a luminosity layer L and a set of chromaticity layers a and b. The color information is present in the a and b layers only. Transforming the blood microscopic images from RGB to CIELAB reduces the color dimension from three (RGB) to two (a and b) and facilitates faster color based image segmentation.

1. Let Irgb represent an original lymphocyte image in RGB color space.
2. Apply L*a*b* color space conversion on Irgb to obtain the L*a*b* image, i.e., Ilab.
3. Construct the input feature vector using a* and b* components of Ilab.

Features can be represented by the space of colour, texture and gray levels, each exploring similarities between pixels of a region. Segmentation refers to the process of partitioning a digital image into multiple regions (sets of pixels). The goal of segmentation is to simplify and change the representation of an image into something that is more meaningful and easier to analyze

2.4 Image Segmentation

Image Segmentation is the process of partitioning a digital image into multiple regions or sets of pixels Actually partitions are different objects in image which have the same texture or color. The result of image segmentation is a set of regions that collectively cover the entire image, or a set of contours extracted from the image. All of the pixels in a region are similar with respect to some characteristic or computed property, such as color, intensity, or texture. Adjacent regions are significantly different with respect to the same characteristics. Edge detection is one of the most frequently used techniques in digital image processing. The boundaries of object surfaces in a scene often lead to oriented localized changes in intensity of an image, called edges. This observation combined with a commonly held belief that edge detection is the first step in image segmentation, has fueled a long search for a good edge detection algorithm to use in image processing. This search has constituted a principal area of research in low level vision and has led to a steady stream of edge detection algorithms published in the image processing journals over the last two decades. Even recently, new edge detection algorithms are published each year. This paper analyses some recent soft computing approaches to detect edges for segmentation.

The term image segmentation refers to the partition of an image into a set of regions that cover it. The goal in many tasks is for the regions to represent meaningful areas of the image, such as the crops, urban areas, and forests of a satellite image. In other analysis tasks, the regions might be sets of border pixels grouped into such structures as line segments and circular arc segments in images of 3D industrial objects. Regions may also be groups of pixels having both a border and a particular shape such as a circle or ellipse or polygon. When the interesting regions do not cover the whole image, we can still talk about segmentation, into foreground regions of interest and background regions to be ignored.

K-Means clustering algorithm is also one of the recent techniques that have been proposed in the area of blood cells analysis. K-Means algorithm is one of the clustering algorithms that classify the input data points into multiple classes based on their minimum distance. In medical imaging, many researchers have proven that K-Means clustering has produced good segmentation image due to its performance in clustering of huge datasets.

2.5 Database Description

We propose a new public and free dataset of microscopic images of blood samples, specifically designed for the evaluation and the comparison of algorithms for segmentation and image classification. The initiative is focused on Acute Lymphoblastic Leukemia (ALL), a serious blood pathology that can being fatal in as little as a few weeks if left untreated, most common in childhood with a peak incidence at 2-5 years of age.

The ALL-IDB2 image files are named with the notation ImXXX_Y.jpg where XXX is a progressive 3-digit integer and Y is a boolean digit equal to 0 if the cell placed in the center of the image is not a blast cell, and equal to 1 if the cell placed in the center of the image is a blast cell. Please note that all images labeled with Y=0 are from for healthy individuals, and all images labeled with Y=1 are from ALL patients.

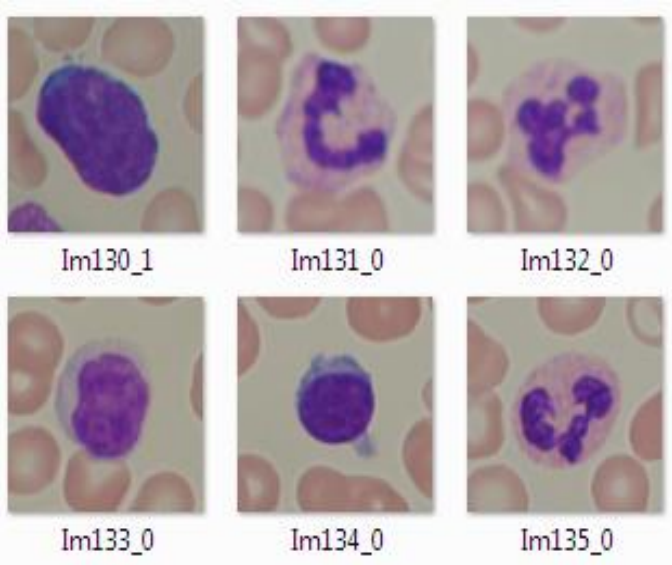

Fig 2.1 y=0 are Healthy

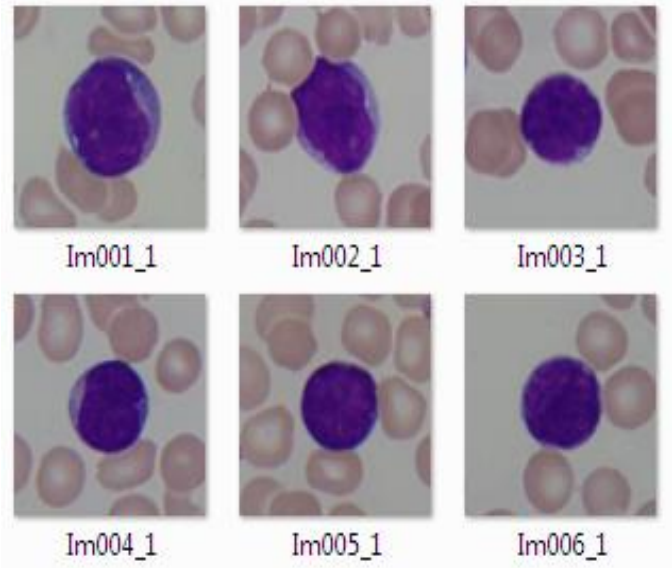

Fig 2.2 y=1 are ALL patients.

- L1: ALL blasts are small and homogeneous. The nuclei are round and regular with little clefting and inconspicuous nucleoli. Cytoplasm is scanty and usually without vacuoles.
- L2: ALL blasts are large and heterogeneous. The nuclei are irregular and often clefted. One or more, usually large nucleoli are present. The volume of cytoplasm is variable, but often abundant and may contain vacuoles.
- L3: ALL blasts are moderate-large in size and homogeneous. The nuclei are regular and round-oval in shape. One or more prominent nucleoli are present. The volume of cytoplasm is moderate and contains prominent vacuoles.

3 Proposed Work

3.1 PSO Image Segmentation

Particle swarm optimization (PSO) is an evolutionary computation technique proposed by Kennedy and Eberhart . The basic idea of PSO is inspired by social behaviour of bird flocking, fish schooling and swarm theory. One of the advantages of PSO is that it is easier to implement and there are very few parameters to adjust. PSO shares many similarities with other computation techniques such as genetic algorithm. PSO has been employed to solve a range of optimization problems, including neural network training and function minimization .Image segmentation methods are a) thresholding, b) edge based segmentation, . Thresholding is the simplest method of image segmentation and separates the pixels of an image into various groups. It works efficiently for bi-modal images. Edge based detection is based on the discontinuity in an image. It is easily effected by the presence of noise and may lead to over as well as under segmentation. Region growing overcomes the drawbacks of early image segmentation techniques.

The swarm is initialized with random particles known as candidate solution and it then searches for optima by updating its position through iterations. Two optimum values define the fitness of objective function first one is the best solution of each particle achieved so far. This value is called as “pbest” solution. Another one is the, best solution tracked by any particle among the whole population. This best value is known as “gbest’ solution $X_i=[X_{j1}, X_{j2, \dots}, X_{jd}]$ and $V_i=[V_{j1}, V_{j2, \dots}, V_{jd}]$. the optimal position of the $_{jth}$ partical the whole swarm, namely the individual optimum and the global optimum are denoted
As $P_i=[p_{j11}, p_{j2, \dots}, p_{jd}]$ and $P_g=[p_{g11}, p_{g2, \dots}, p_{g,d}]$
Respectively individual or partical swarm update their velocities and positions according to the following formulas:

$$Vi(t + 1)=w\ Vi(t) + c1.r1\ (Pbest,i(t) - Xi(t)) + c2.r2\ (Gbest,i(t) - Xi(t))$$

$$Xi(t + 1) = Xi(t) + Vi(t + 1)$$

Where Xi(t) , Vi(t) indicate the position the velocity of particle . Pbest,i indicate the personal best position of particle. Gbest,I indicates the global best position achieved so far. c1 and c2 position acceleration constant r1 and r2 are random values generated between [0, 1]. w is inertia weight used to provide balance between local and global search
where the inertia weight coefficient *w* indicates the ability to track the previous speed; the acceleration coefficients *c*1and *c*2 are used coordinate the degrees of tracking the individual and global optimum; and *r*1 and *r*2 are two random numbers drawn from the uniform distribution on interval(0, 1).

The update equation of the velocity consists of the previous velocity component, a cognitive component and a social component. They mainly controlled by three parameters the inertia weight and two acceleration coefficients .From the theoretical analysis the trajectory of a PSO algorithm, the trajectory of a particle *xi* converges to a weighted mean of *Pi* and *Pg*. Whenever the particle converges ,it will “fly” to the individual best position and the global best position (Clerc and Kennedy 2002). According to the update equation, the individual best position of the particle will gradually move closer to the global best position. Therefore, all the particles will converge onto the global best position.

The segmentation is based on the measurements taken from the image and might be greylevel, colour, texture, depth or motion. Image segmentation techniques are categorized into three classes: Clustering, edge detection, region growing. Some popular clustering algorithms like k-means are often used in image segmentation adjacent regions are significantly different with respect to the same characteristic(s).

Segmentation is mainly used in medical imaging, Face recognition, Fingerprint recognition, Traffic control systems. Minimum cross entropy thresholding method is very time consuming in multilevel thresholding as compared to bi-level thresholding for complex image segmentation.

3.2 Genetic Algorithm

Genetic algorithms (GA) are adaptive heuristic search algorithm based on the evolutionary ideas of natural selection and genetic. Genetic algorithm is a method for moving from one population of "chromosomes" to a new population by using a kind of "natural selection" together with the genetic inspired operators of crossover, mutation and inversion. Each chromosome represents a solution of the problem. In a search space, best of them are selected from the solution set available in search space.

After determining the population size and manner of encoding, each solution or chromosome is evaluated. To do this fitness function is used. Fitness function depends on our problem. Based on this fitness function, fitness value is calculated for each chromosome. This fitness value tells us that how close the solution is to solve a particular problem. After the fitness function is determined for each member of the population, genetic operators should be applied on them to prevent premature

The Genetic algorithm is a model of machine learning which derives its behavior from a metaphor of the processes of evolution in nature. This is done by the creation within a machine of a population of individuals represented by chromosomes, in essence a set of character strings that are analogous to the base-4 chromosomes that we see in our own DNA. The individuals in the population then go through a process of evolution.

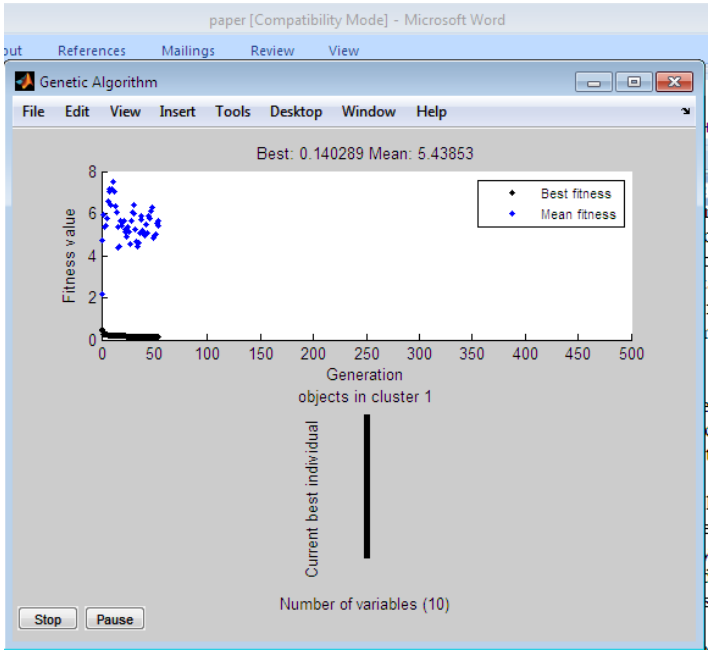

Fig 4.2 represents the Best Mean value of leukemia images

4.Experiments and Result

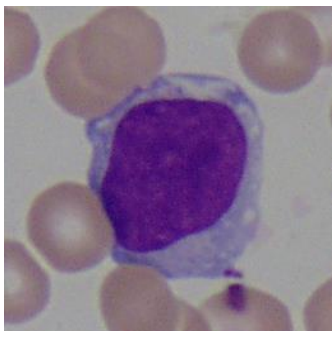

Fig 4.1 original image

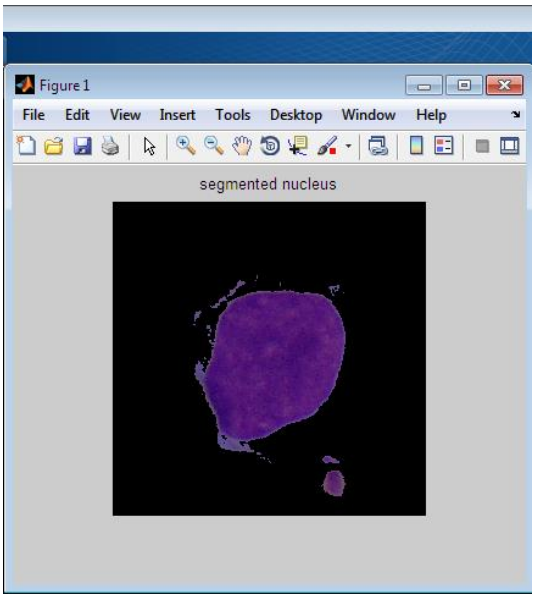

Fig 4.3 Segmented nucleus

4. Conclusions

This paper has demonstrated a proposed framework for segmenting white blood cells using integration of concepts in digital image processing. The proposed scheme has two parts: The nucleus segmentation part is based on morphological analysis, genetic pso and the cyto- plasm segmentation is based on pixel-intensity thresholding. The results show that the proposed method is able to yield 92% accuracy for nucleus segmentation and 78% for cytoplasm segmentation.

References

[1] Greco, S, B. Matarazzo And R. Slowinski,"Rough Approximation Of Preference Relation By Dominance Relations". Ics Research Report 16/96, Warsaw University of Technology and European Journal of Operational Research, pp: 63-83, 1999.

[2] Greco, S., B. Matarazzo and R. Slowinski, "A New Rough Set Approach to Evaluation of Bankruptcy Risk". In Operational Tools In The Management Of Financial Risk, pp: 121-136

[3] Greco S., Matarazzo B., Slowinski R., The use of rough sets and fuzzy sets in MCDM. In T. Gal, T. Stewart and T. Hanne (eds.) Advances in Multiple Criteria Decision Making, chapter 14, Kluwer Academic Publishers, Boston, 1999, pp:14.1-14.59

[4] Greco, S., B. Matarazzo, R. Slowinski and J. Stefanowski. "An Algorithm for Induction Of Decision Rules Consistent With The Dominance Principle". Proc. Rsctc'Conference, 2000

[5] Gomes, P.R. ; Soares, F.O. ; Correia, J.H. ; Lima, C.S. "ECG Data-Acquisition And Classification System By Using Wavelet-Domain Hidden Markov Models" Engineering In Medicine And Biology Society (Embc), 2010 , pp: 4670 – 4675

[6] Jewajinda,Y.;Chongstitvatana,P."Fpga-Based Online-Learning Using Parallel Genetic Algorithm And Neural Network For ECG Signal Classification" Electrical Engineering/Electronics Computer Telecommunications And Information Technology (Ecti-Con), 2010 International Conference On publication Year: 2010 , pp: 1050 – 1054

[7] Hyejung Kim ; Yazicioglu, R.F. ; Merken, P. ; Van Hoof, C. ; Hoi-Jun Yooecg "Signal Compression And Classification Algorithm With Quad Level Vector For ECG Holter System" Information Technology In Biomedicine, IEEE Transactions On Volume: 14 , Issue: 1 2010 , pp: 93 – 100

[8] Ming-Yao Yang ; Wei-Chih Hu ; Liang-Yu Shyu , "ECG Events Detection And Classification Using Wavelet And Neural Networks" Engineering In Medicine And Biology Society", Proceedings Of The 19th Annual International Conference Of The IEEE,1997,pp:280-281

[9] Adib, A. ; Haque, M.A. "ECG Beat Classification Using Discrete Wavelet Coefficients Health Informatics And Bioinformatics "(Hibit), 2010 5th International Symposium On 2010 , pp: 1 - 6

[10] Can Ye ; Coimbra, M.T. ; Kumar, B.V.K.V. arrhythmia Detection And Classification Using Morphological And Dynamic Features Of ECG Signals Engineering In Medicine And Biology Society (Embc), 2010 Annual International Conference Of The IEEE 2010 , pp: 1918 – 1921

[11] Gomes, P.R. ; Soares, F.O. ; Correia, J.H. ; Lima, C.S. "Ecg Data-Acquisition And Classification System By Using Wavelet-Domain Hidden Markov Models" Engineering In Medicine And Biology Society (Embc), 2010 , pp: 4670 – 46

[12] Patil, P.B. ; Chavan, M.S,"A Wavelet Based Method for Denoising Of Biomedical Signal Pattern Recognition ", Informatics And Medical Engineering (Prime), 2012,pp:278-283