

SURVEY ON CHROMOSOME IMAGE ANALYSIS FOR ABNORMALITY DETECTION IN LEUKEMIAS

Vinya Vijayan¹, Remya R S², Sabeena K³

¹PG Scholar, Computer Science and Engineering, College of Engineering Karunagappally, Kerala, Country

²Assistant Professor, Computer Science and Engineering, College of Engineering Karunagappally, Kerala, Country

³Assistant Professor, Computer Science and Engineering, College of Engineering Karunagappally, Kerala, Country

Abstract

Development of an automated system for the chromosome classification and abnormality detection in the case of Leukemias has significant importance in the field of cytogenetics. Such systems can be effectively used to find the genetic disorders and used in treatment evaluation procedures, prognosis prediction and in follow up treatment with the help of karyotyping. This is possible because the leukemia's are generally characterized by numerical and structural chromosomal abnormalities. During the last three decades, several developments in the field of automated chromosome analysis have been occurred. Different preprocessing, segmentation, feature extraction and classification methods are applied to develop an automatic karyotyping system. But still now, the classification and the abnormality detection of chromosomes are done in the laboratories with the help of human intervention which is time consuming and labor intensive. Further research is needed in this field to develop completely automated abnormality detection systems which are capable of effectively identify the abnormalities without the human intervention. Here we have done a survey on different methods of karyotyping existing so far that leads to efficient systems for abnormality detection.

Keywords: Abnormality Detection, Chromosomes, Leukemia and Karyotyping.

1. INTRODUCTION

Chromosomes are the double helical strands in the cell and they carry the genetic information carriers known as genes. Therefore the genetic disorders like leukemias are reflected in chromosomal abnormalities and can be found using the chromosome image analysis procedures. There are 23 pairs of chromosomes in a normal metaphase cell, in which the first 22 pairs are called the autosomes and the remaining one pair is sex chromosomes. Different types of Leukemia's can be confirmed by identifying different types of chromosomal abnormalities.

The flow chart of the automated systems for karyotyping and abnormality detection is shown in Fig-1. The microscopic images of the G-banded metaphase spreads are captured using the camera attached to the microscope. The quality of the automated abnormality detection systems depends on the resolution and clarity of the obtained chromosome images. To improve the quality of the chromosome images different preprocessing techniques such as noise removal and contrast enhancement techniques are applied on the captured image. The individual chromosomes from the captured metaphase images are segmented using different segmentation techniques. Medial axis segmentation is essential because many of the essential features are extracted from the approximate medial axis of the chromosome.

The desired features are extracted from the segmented chromosomes for automated karyotyping and abnormality detection. Feature selection methods are applied to reduce

the dimensionality of the feature set. The classifier, based on the computed features classifies the chromosomes into different classes and the karyotype is formed. The different chromosomal abnormalities are detected from the further analysis of karyotypes. The image of the metaphase spread and is shown in Fig-2 and its karyotype is shown in Fig-3.

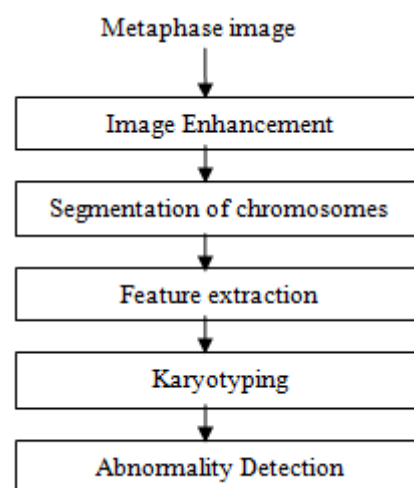


Fig -1: Flow chart of the abnormality detection systems.

The remainder of this paper is outlined as follows. Section 2 is a brief description of the works and their contributions in the image enhancement. Section 3 lists various segmentation procedures and section 4 describes different feature extraction methods commonly used in the abnormality detection systems. The classifiers for the karyotyping are

described in section 5. Section 6 focuses on the abnormality detection methods and in section 7 the findings from the survey are presented.



Fig -2: Microscopic image of the metaphase spread.

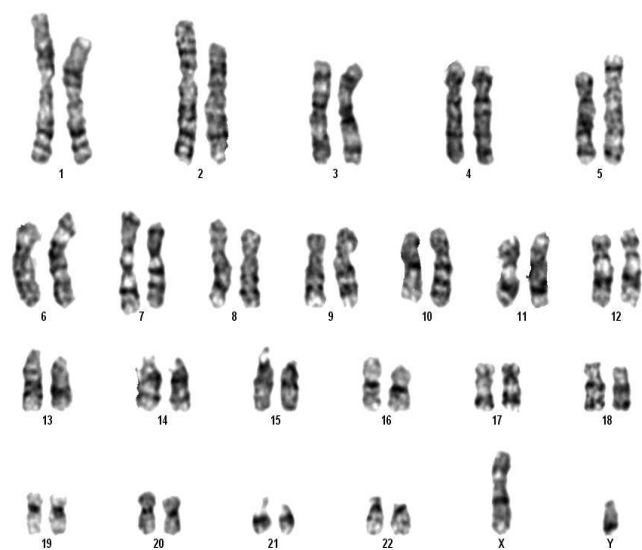


Fig -3: Karyotype of the metaphase spread shown in Fig -2.

2. IMAGE ENHANCEMENT

Since 1980s, researchers tried to enhance the metaphase spreads and the chromosome images, in the hope that the image enhancement techniques could lead to better and effective segmentation and classification of chromosomes. Over all, the performance of the automated karyotyping systems and the abnormality detection systems can be improved using the enhancement techniques. The purpose of the enhancement techniques are noise removal, contrast enhancement, separation of overlapping and occluded chromosomes and the geometrical corrections to nullify the bending effects.

2.1 Noise Removal

The images with noise are difficult to process and the spotting of the bands becomes unattainable. Suppression of the noise from the low quality images is desirable before the segmentation and classification. The traditional smoothing filters such as Gaussian lowpass and median filters are used for reducing the random noise in [1].

2.2 Contrast Enhancement

Contrast is the difference between the intensity of a pixel and its neighbors and different contrast enhancement techniques are applied on the metaphase spread.

Histogram equalization is used to improve the contrast and image quality in [2],[3]. Spatially enhanced images contain inhomogeneties and histogram equalization is used to overcome these irregularities irrespective of the geometrical distortions or the size or shape of the chromosomes. Otsu's automatic thresholding algorithm was used to calculate the optimized threshold of the particular chromosome images in [4]. Threshold optimized by fuzzy compactness gives better enhancement of the images [5].

2.3 Geometric Corrections

The curved chromosome images cause difficulties in the feature extraction. Therefore geometrical corrections need to be applied on the chromosome images to nullify the bending effects.

An efficient straightening algorithm is used in [6] based on the horizontal and vertical projection vectors of the binary image. Highly curved chromosomes can be straightened using this method. Another three step algorithm that uses the seed region extraction and straightening of the chromosome axis is suggested in [7].

3. SEGMENTATION

Thresholding is one of the popularly used chromosome segmentation method and it is the best one in the absence of overlapping chromosomes [4]. The individual chromosomes were segmented by a suitable gray-level threshold found from the density histogram. The thresholding technique was later used by many researchers for the chromosome isolation [2],[8]. This procedure suited well for the individual chromosome separation from the karyotypes. Because of the limitations of the thresholding several new methods are investigated by the researchers. But the segmentation by thresholding did not work well for the touching and overlapping chromosomes and leave these types of chromosomes as clusters and the produced clusters have to be separated manually.

The segmenting efficiency of different types of applied thresholds such as kapur's threshold, k-means clustering of algebraic moments, fuzzy c-means clustering on intensities (FCM), Local re-thresholding (LRT), Multi stage adaptive thresholding (MAT), Multi-Thresholding with Particle Swarm and NelderMead Optimization (PSO-NM) and Region based level sets (RBLS) are summarized in [9].

Multichannel watershed-based segmentation is applied in the multispectral images of chromosomes for segmentation in [10]. Later watershed transform was used for the normal G-banded chromosome images for isolation [2]. The watershed transform have some additional advantages over other segmentation methods. The watershed lines form closed and connected regions and they always corresponds to obvious contours of objects. The main disadvantage of this technique is the over-segmentation and there were several methods [10] to avoid these drawbacks by the use of preprocessing or post-processing.

3.1 Touching and Overlapping Chromosomes

The segmentation of touching and overlapping chromosomes is important in the abnormality detection because the occluded chromosome segments may contain the bands and other useful information; and in the absence of that information there arise difficulties in abnormality detection.

The touching and overlapped images are separated using image contour and hypothesis construction in [11]. Algorithm based on the cut points selection and k-means clustering is applied for disentangling overlapping chromosomes in [12]. Another algorithm based on the cut points and cut lines from the detected contours is applied in [13].

3.2 Medial Axis Segmentation

Medial axis or the chromosome axis is a longitudinal axis, about which the sister chromatids are symmetrical and medial axis extraction is important because it provide many useful geometric descriptions about the chromosomes. The chromosome axis is used an effective tool for the feature extraction purpose because it is invariant of the geometrical changes or the bending.

Poor mans skeleton (PMS) is a non-parametric technique suggested for the medial axis extraction in [1]. However, the PMS cannot be applied in the case of bended chromosomes and Hilditch's algorithm [1] is suggested in that case. The main drawback is that, PMS only produce an approximation of the chromosomal axis and it is not an exact one. Furthermore, the studies showed that PMS works well with long chromosomes, but cannot be applied to short chromosomes.

Another method for obtaining the chromosome axis is the iterative morphological thinning algorithm [14]. The thinning aims to produce a 1-pixel wide curve form the boundary of the chromosome object. The bifurcations produced during the thinning are deleted using pruning and later pixel sorting and curve smoothing techniques were employed. Axis extension at the two chromosome ends was done using a simple polynomial extrapolation procedure. The thinning produced an accurate approximation to the medial axis and is better than the PMS. Thinning can be used for the medial axis extraction of the long and bended chromosomes.

The combination of the thinning algorithm with the parametric techniques such as curve fitting can be used for effective medial axis modeling [15]. The boundary extraction followed by thinning was applied and later curve fitting or regression analysis was used to produce a best fit of the medial axis. The data was fitted to the cubic polynomial by the least square curve fitting. The major advantage of this approach is that, there is no need for the further extension of the chromosome axis tips. Also this method works well for the bended and long chromosomes.

4. FEATURE EXTRACTION

The outcome of the abnormality detection process depends on the quality of the extracted features. An optimal and small feature set is one of the key factors determining the efficiency of abnormality detection. Selecting the optimal features of choice is a challenging task and different feature selection processes are applied on the feature set to produce an optimal feature vector.

Knock-out algorithm which is a backward technique and a scatter criterion which depends on the within-class scatter matrix are employed in [16] for feature selection. The knockout algorithm reduces the 66- dimensional features into 1-10 pattern vectors. Popular dimensionality reduction methods such as Linear Discriminant Analysis (LDA), Discrete Cosine Transform (DCT) and Principal Component Analysis (PCA) are compared in [17] for the same feature set and among these LDA gives the best performance.

4.1 Length Related Features

Length of a single chromosome is an explicit distinguisher of the chromosomes within a cell due to the fact that the chromosome length gradually declines from class 1 to 22 except the class 21. Furthermore, the shortening or lengthening of a chromosome in a cell is an explicit indicator of abnormalities such as deletion, duplications and translocations. Therefore, the length related features play a crucial role in chromosome classification and abnormality detection.

But the length of the chromosomes cannot be used as an effective feature set due to the low image quality of the metaphase spreads. The length related features are obtained from the segmented medial axis of the chromosome. Therefore the accuracy of the chromosome length depends on the quality of the medial axis transformation.

The length of a single chromosome can be calculated from the extracted medial axis of the corresponding chromosome [4]. Relative Length (RL) is the length of the i th chromosome to the total length of all the chromosomes present in the cell [17],[18]. Centromeric Index (CI) is the ratio of the length of the short arm (p) to the whole length of the chromosomes [4],[17],[18].

4.2 Shape Related Features

The chromosome area of each segmented chromosome is obtained by summing the area of each individual pixel in the chromosome [1],[3]. The perimeter of the chromosome is obtained through chain codes using the direction and length information [3].

4.3 Banding Patterns

Unique banding pattern on the chromosomes play a crucial role in automated the chromosome classification and abnormality detection systems.

Piper et al. [4] suggested band profile extraction from the approximated medial axis for calculating the weighted density profile features. It denotes a graph of some property of the chromosome sampled at a sequence of points along the medial axis. The density profiles [DP] represent the average gray level values along the perpendicular lines to the medial axis of the chromosome [2],[3],[13],[19]. The band chromosome area of each bands are identified using thresholding technique [3]. The area of bands can be compared with the standard ideograms of the same resolution to detect the abnormalities.

4.4 Surface Features

Wavelet Transform (WT) with the Daubechies wavelet basis in four decomposition levels is calculated for the DP of the chromosome in [17]. For better feature vector creation from the WT features, Linear Discriminant Analysis (LDA) which is the commonly applying dimensionality reducing method is applied on the extracted features. The 2-D Discrete Wavelet Transform (DWT) is applied on the gray scale images to extract the frequency and location related information such as standard deviation, covariance and bag [2].

4.5 Numerical Features

The features that depend on the number of chromosomes have an impact on the numerical abnormality detection. DNA index is the ratio of the average number of chromosomes in the abnormal metaphase spread of a patient to the normal chromosome count in a meta-phase spread [22]. DNA index can be used as an effective feature for numerical abnormality detection [22].

5. KARYOTYPING

Karyotyping is the layout of the chromosomes in the decreasing order of their length except the sex chromosomes. Technically karyotyping is the classification of chromosomes into pairs in the decreasing order of their length. Different types of classifiers are employed for automated karyotyping of chromosomes.

Artificial Neural Networks (ANN) is widely used for chromosome classification in recent studies. The two layer ANN is used for the classification of Q-banded chromosomes [8]. The three layer feed-forward perceptron

neural network trained with the back propagation algorithm is used for classification in [17]. The number of input nodes depends on the dimension of the feature vector used and the numbers of output nodes are equal to the chromosome classes to be separated. The classification efficiency was 99.3% with the WT based features along with RL and CI. However over fitting of the training data is a major challenge in ANNs.

Probabilistic neural network with two layers are used for chromosome classification in [3]. The classification accuracy of the network is 68.19% for female and 61.30% for male [3]. This classifier need less training time comparable to the back propagation neural network. Furthermore, it has the added advantage that the incremental learning is possible which allows the addition of training data without retraining of the network.

Different distance classifiers can also be applied for classification purposes. K-Nearest Neighbor Classifier (k-NN) is the simplest classification algorithm in which the objects are classified on the basis of their neighbors. k - NN is used for the pairing of chromosomes in [2]. A two stage classification scheme which combines the context aware classifier and the support vector machines (SVM) can also be applied for classification [14].

Even though different classifiers are exists for automated karyotyping, the efficiency of these classifiers on the varying metaphase spreads is still a challenging task. Further research is needed to increase the accuracy of the automated karyotyping systems.

6. ABNORMALITY DETECTION

The abnormalities can be either structural or numerical and the automated abnormality detection systems have significant importance in the prognosis and treatment evaluation procedures [23]. The different chromosomal abnormalities are summarized by the various researchers [22] and the automated systems tried to find these specific abnormalities to confirm the particular diseases.

6.1 Numerical Abnormality Detection

The numerical abnormalities are characterized by a change in chromosome number within a cell; and the total number of chromosomes in a numerically abnormal cell is greater than or less than the normal chromosome count [23]. Based on the chromosome count the karyotype can be classified into different ploidy groups such as Hypodiploidy, Pseudodiploidy, Hyperdiploidy and Tetraploidy [22], [23].

Numerical abnormality detection for Acute Lymphocytic Leukemia (ALL) from the metaphase spreads is done in [22]. In that method, connected component labeling is used to count the number of chromosomes and then DNA index is calculated for each metaphase spread sample to detect the ploidy group. Based on the ploidy group the prognoses are predicted for each patient dataset.

6.2 Structural Abnormality Detection

Structural abnormalities denote the variation from the normal chromosome structures [23] and the genetic disorders; especially the leukemias are generally associated with the structural abnormalities. The abnormality can be deletion, duplication, inversion or the commonly occurring translocations. Deletion is the loss of chromosome segments and duplication is the addition of chromosome segments. In inversion, the chromosome segments are inverted and in translocation is the chromosome segments are exchanged between the nonhomologous chromosomes.

Each type of Leukemia has specific structural variations and sometimes, the abnormalities are the key indicators of a particular type of Leukemia. For example, the translocation t(9;22) is an indicator of either ALL or Chronic Myeloid Leukemia (CML). By analyzing the chromosome images, the cytogenetic expert can manually find the abnormality which is the current clinical practice. However, some attempts are made to develop an automated structural abnormality detection system.

Xingwi et al. [24] proposed an automated scheme for detecting the Chronic Myeloid Leukemia (CML) from the metaphase spreads. The idea underlying in the development of the system is that the presence of only one normal chromosome 22 in the cell indicate the presence of the translocation t(9;22) which is mostly associated with CML. The method searches for a normal chromosome 22 in the sample through an adaptive template matching algorithm. The correlation based similarity measure Normalized Cross correlation (NCC) is used for the matching process. Pearson's correlation coefficient is calculated between the reference chromosome 22 and the candidate chromosome; and the candidate with the highest similarity score is selected as the chromosome 22. After obtaining one chromosome 22, the algorithm searches for the second chromosome 22 and the absence of a second chromosome 22 in the sample indicate the presence of CML. The major drawback of the scheme is that, the absence of chromosome 22 is not only the indicator of CML but also the indicator of other Leukemia's.

Another method for detecting the abnormal chromosome segments involved in the translocations of solid tumors through the banding analysis of the spectral karyotyped image is discussed in [25]. In this method, the density profiles of the reference and abnormal chromosomes are compared using the Dynamic Time Warping (DTW) algorithm to detect the Break Points (BP) and the Recombination Points (RP) of the abnormal sections. The DTW is a signal comparison algorithm and in which the DTWCost between the DP of the reference template and the abnormal chromosome are computed to find the BPs.

Only limited works are exists for the automated abnormality detection and therefore further contributions are required in this field to develop efficient solutions.

7. CONCLUSION

The chromosomal abnormality detection of Leukemias from the metaphase spreads is a challenging task. Different preprocessing, segmentation and feature extraction methods are suggested by various researchers in the last three decades for the classification of chromosomes.

Even though different classification methods are exist, still the karyotyping is performed by the automated karyotyping systems with the help of human intervention in the clinical laboratories. Therefore, new methods are required to develop fully automated karyotyping systems.

Currently, the abnormality detection is done through the manual analysis of chromosome images by the cytogenetic expert. Mimicking the actions of the cytogenetic expert is a major technical problem in the automated abnormality detection systems. Abnormality detection has great role in the prognosis prediction and follow up treatment. Further research is needed to develop an automated system for the detection of chromosomal abnormalities in various Leukemias.

ACKNOWLEDGEMENTS

We would like to thank all the reviewers for their valuable comments.

REFERENCES

- [1]. A. Weeks, H. Myler, and G. Wenaas, "Computer-generated noise images for the evaluation of image processing algorithms," *Opt. Eng.*, 1993.
- [2]. S. Janani, R. Nandakumar, and M. Nirmala, "Feature extraction and pairing of g-band chromosome images using k-nearest neighbour classifier," *International Journal of Computer Science And Technology*, vol. 3, no. 2, pp. 137-140, Apr-Jun 2012.
- [3]. S. Rungruangbaiyok and P. Phukpattaranont, "Chromosome image classification using a two-step probabilistic neural network," *Songklanakarin J. Sci. Technol.*, vol. 32, no. 3, pp. 255-262, May-Jun 2010.
- [4]. J. Piper and E. Granum, "On fully automatic feature measurement for banded chromosome classification," *Cytometry*, vol. 10, pp. 242-255, 1989.
- [5]. S. K. Pal and A. Rosenfeld, "Image enhancement and thresholding by optimization of fuzzy compactness," *Pattern Recognition Letters - PRL*, pp. 77-86, Feb. 1988.
- [6]. M. J. Roshtkhari and S. K. Setarehdan, "Novel algorithm for straightening highly curved images of human chromosome," *Pattern Recognition Letters - PRL*, vol. 29, no. 9, pp. 1208-1217, 2008.
- [7]. S. Khan, A. DSouza, J. Sanches, and R. Ventura, "Geometric correction of deformed chromosomes for automatic karyotyping," *IEEE EMBS*, pp. 4438-4441, Aug. 2012.
- [8]. E. Poletti, E. Grisan, and A. Ruggeri, "A modular framework for the automatic classification of chromosomes in q-band images," *Computer Methods and Programs in Biomedicine*, pp. 120-130, 2012.

- [9]. E. Poletti, F. Zappelli, A. Ruggeri, and E. Grisan, "A review of thresholding strategies applied to human chromosome segmentation," *Computer methods and programs in biomedicine*, pp. 679–688, 2012.
- [10]. P. S. Karvelis, A. T. Tzallas, D. I. Fotiadis, and I. Georgiou, "A multichannel watershed-based segmentation method for multispectral chromosome classification," *IEEE Transactions on Medical Imaging*, vol. 27, no. 5, pp. 697–707, May 2008.
- [11]. M. V. Munot, M. Joshi, N. Sharma, and G. Ahuja, "Automated detection of cut-points for disentangling overlapping chromosomes," *IEEE Point-of-Care Healthcare Technologies (PHT)*, pp. 120–123, Jan. 2013.
- [12]. D. Somasundaram and V. V. Kumar, "Separation of overlapped chromosomes and pairing of similar chromosomes for karyotyping analysis," *Elsevier*, pp. 274–281, 2014.
- [13]. N. Madian and K. B. Jayanthi, "Overlapped chromosome segmentation and separation of touching chromosome for automated chromosome classification," in *IEEE EMBS*, vol. 34, Aug-Sep. 2012, pp. 5392–5395.
- [14]. C. Markou, C. Maramis, A. Delopoulos, C. Daiou, and A. Lambropoulos, "Automatic chromosome classification using support vector machines."
- [15]. X. Yanga, W. Wua, and C. C. Tsengb, "Algorithms for modeling structural changes in human chromosomes," *Computer Methods and Programs in Biomedicine*, pp. 171–182, 2013.
- [16]. B. Lerner, "Toward a completely automatic neural-network-based human chromosome analysis," *IEEE Transactions On Systems, Man, And Cybernetics Part B: Cybernetics*, vol. 28, no. 4, pp. 544–551, Aug 1998.
- [17]. M. J. Roshtkhari and S. K. Setarehdan, "Linear discriminant analysis of the wavelet domain features for automatic classification of human chromosomes," in *ICSP*, 2008, pp. 849–852.
- [18]. S. Y. Ryu, J. M. Cho, and S. H. Woo, "A study for the feature selection to identify giemsa-stained human chromosomes based on artificial neural network," in *Annual EMBS Int. Conf.*, vol. 23, 2001, pp. 691–692.
- [19]. S. G. Palalica and M. Canb, "Extracting gray level profiles of human chromosomes by curve fitting," *Southeast Europe Journal of Soft Computing*, pp. 66–71, 2008.
- [20]. A. S. Arachchige, J. Samarabandu, J. Knoll, W. Khan, and P. Rogan, "An image processing algorithm for accurate extraction of the centerline from human metaphase chromosomes," in *IEEE 17th International Conference on Image Processing*, 2010, pp. 3613–3616.
- [21]. A. D. Ward and G. Hamarneh, "The groupwise medial axis transform for fuzzy skeletonization and pruning," *IEEE Transactions on Pattern Analysis And Machine Intelligence*, pp. 1–14, 2009.
- [22]. X. Wang, S. Li, H. Liu, J. J. Mulvihill, W. Chen, and B. Zheng, "A computer-aided method to expedite the evaluation of prognosis for childhood acute lymphoblastic leukemia," *Technology in Cancer Research and Treatment*, vol. 5, no. 4, pp. 429–436, Aug. 2006.
- [23]. C. H. Pui, W. M. Crist, and A. T. Look, "Biology and clinical significance of cytogenetic abnormalities in childhood acute lymphoblastic leukemia," *The American*

Society of Hematology, vol. 76, no. 8, pp. 1449–1463, october 1990.

[24]. B. Legrand, C. S. Chang, S. H. Ong, S.-Y. Neo, and N. Palanisamy, "Automated identification of chromosome segments involved in translocations by combining spectral karyotyping and banding analysis," *IEEE Transactions On Systems, Man, And Cybernetics Part A: Systems And Humans*, vol. 38, no. 6, pp. 1374–1384, Nov 2008.

[25]. X. Wang, B. Zheng, S. Li, J. J. Mulvihill, X. Chen, and H. Liu, "Automated identification of abnormal metaphase chromosome cells for the detection of chronic myeloid leukemia using microscopic images," *Journal of Biomedical Optics*, vol. 15, no. 4, Jul 2010.

BIOGRAPHIES



Vinya Vijayan received the B.Tech. degree in Computer Science and Engineering from Sree Buddha College of Engineering in 2013. She is now doing M.Tech in Computer Science and Engineering with specialization in Image Processing in College of Engineering Karunagappally.



Remya R S received the B.Tech. in Computer Science and Engineering from CUSAT in 2006 and M.Tech. in Computer and Information Technology from M. S University in 2010. Research areas are Digital Image Processing, specifically focus on Video Processing and Medical Imaging. She now with College of Engineering Karunagappally as an Assistant Professor.



Sabeena K received the B.Tech. degree in Computer Science and Engineering Kannur University in 2004. She now with College of Engineering Karunagappally as an assistant Professor in the department of Computer Science and doing M.Tech in Computer Science and Engineering in College of Engineering Trivandrum. Research areas are medical imaging and machine learning.