Texture based Leaf Disease classification using Machine Learning Techniques

Shravankumat Arjunagi, Nagaraj B. Patil

Abstract- Machine learning techniques has emerged as a potential field in many of present day agricultural applications. One of these applications is the identification and classification of leaf diseases. In this paper, a triangular based and OTSU based methods are applied for segmentation, Textural features primarily based on GLCM are obtained for these segmented images using k-means clustering technique, further classification of different leaf disease is performed using an SVM based classification. The proposed method resulted in an overall classification accuracy of 70% using the triangular based segmentation with an AUC of 0.63.

Keywords: Machinelearning, Triangular, Segmentation, SVM.

I. INTRODUCTION

The Indian economy derives about 70 % of its revenue from agriculture. The quality and quantity of agricultural products are usually affected by diseases pertaining to various crops. These disease change vital functions of the crop such as photosynthesis, pollinations, transpiration, germination and fertilization. Some of these diseases are caused due to infection from pathogens, fungi, virus and bacteria. Conventional methods of detecting such diseases involved continuous observations from farmers or by agricultural experts, however, these techniques are often time consuming, expensive and requires additional manpower.

With recent developments in the field of machine learning and computer automation, many applications such as healthcare, military, commercial industry has harnessed its potential, This includes even in the field of agriculture and related imaging applications. Leafs are often vulnerable to diseases during their early stages, Therefore, it is essential to recognize and diagnose these diseases in the initial stages. (Gavhale and Ujwalla, 2014; Rumpf, 2010). This is possible by using machine learning techniques as these diseases could be characterized by features such as texture, color and shape of the diseased leaf (Al-Hiary et al., 2011; Gulhane and Gurjar, 2011; Patil and Kumar, 2011).

Many techniques were implemented in identifying and classifying diseases in leaf images.

Some of the works include the use of Gabor filter, homogenous pixel counting, fuzzy C-means and Artificial Neural Networks (ANN) for extracting features, segmentation and classifying of leaf diseases (Kulkarni and AshwinPatil, 2012; Revathi and Hemalatha, 2012; Mokhled, 2013; Kanjalkar and Lokhande, 2014).More recent methods on imaging of leaf disease include Content Based Image Retrieval

(CBIR) which uses RGB based color histogram (Chakravarti and Meng, 2009), Gray Level Co-occurrence Matrix (GLCM) using K-means (Rasli, 2012), Color moments (Redi, 2011, Weng, 2013) and local binary patterns for texture analysis (Doshi and Schaefer, 2013). Edge detection methods such as Shape Adaptive Discrete Cosine Transform (SA-DCT) (Belloulata, 2014), Scale Invariant Feature Transform (SIFT) with Canny Edge Detection (CED) (Bandaru and Naik, 2014) were used to identify features defining spatial characteristics.

However, these methods do not consider various factors such as orientation and scaling of images, sensitivity to illuminations, correction to noise factors among others which affects imaging features such as shape, size, textures, etc. These variations further cause alterations during segmenting of leaf images. Other issues pertaining to classification involves feature redundancy, dimensionality reduction, bias correction, image quality assessment during segmentation, etc.However, recent developments in imaging have led in identifying imaging features characteristic to particular type of fungi. Therefore, In this paper, identifying leaf disease characteristics involving Alternata-Alternata and Anthracnose is performed, which in turn could distinguish between these two regions using machine learning techniques.

The paper is structured as follows, the second section involves reviewing of literature of various methods of segmentation, classification and validation techniques in extracting and diagnosing image features in leaf diseases. The third section describes the functionality and implementation of the proposed method, followed by results and discussion respectively, with a section concluding the overall experimental details with findings.

II. REVIEW OF LITERATURE

Various imaging and analytical techniques has evolved over the years in identifying and diagnosing diseases in leaf. The emerging trend of automation has further increased the research interest in the field of agriculture and related applications. A review of these methods is being illustrated in this section.

Methods of texture based analysis for deriving imaging characteristics of leaf diseases were proposed by works of Chakravarti, 2009; Rasli, 2012, Redi, 2011; Weng, 2013 and Doshi and Schaefer, 2013.

Secondly, segmentation plays an important role in the diagnosis of leaf diseases.

Revised Manuscript Received on October 15, 2019

Mr. Shravankumar Arjunagi, Assistant professor in Department of Computer Science and Engineering at APPA Institute of Engineering and Technology Kalaburgi- Karnataka.

Dr. Nagaraj B. Patil, Associate professor and HOD Dept. of CSE & ISE at Government College of Engineering, Raichur Karnataka

Conventional methods of segmentation involves thresholding methods such as adaptive thresholding , histogram based thresholding (OTSU), etc. Other methods involves segmentation through morphological methods such as rectangular method, triangle method, square method, etc. Recently, methods involves the use of analytical methods such as k-means based clustering, KNN based methods, etc.

Ma, 2009 performs a review on various segmentation techniques and broadly classifies into three categories involving threshold based, pattern based and deformable based models. Some of the main objectives of these models involves identifying and segmenting the tissues in pelvic cavity area.

Tavares, 2009 describes challenges faced during computational analysis mainly involving tasks pertaining to segmentation, representative feature extraction, matching, alignment, tracking, motion analysis and estimating the deformation.

Sanjay.B, 2013 proposed a vision based detection algorithm with masking of green pixels and color co-occurrence method. However, the method had limitation on the classification performance which could be addressed through neural networks.

Mrunalini. R, 2012 proposed a k-means based clustering method to perform pattern recognition on crop diseases with neural networks. This method can significantly provide an accurate detection the in case of stem and root diseases. However, the method has limitation of implementation of soft computing techniques which could be used to classify various diseases in leaf.

S. Arivazhagan, 2013 proposed a classification method for leaf diseases based on textural features such as color co-occurrence matrix with SVM based classification. However, the issue of sample size could be improved. The method also had a limitation of involving other aspects of imaging feature characteristics such as shape and color for disease identification.

Anand. H Kulkarni, 2012 proposed a gabor based filtering approach to with a classifier using Artificial Neural Network (ANN). This yielded a significant performance of the classification with a classification rate of 91%. However, the method had limitation on the rate of recognition with respect to leaf diseases.

Sabah Bashir, 2012 proposed a texture based segmentation method on a plant type MalusDomesticausing a combination of co-occurrence matrix and k-means based clustering method. Feature dependency and other aspects could be further observed using Bayes classifier, Kmeans and Principal Component Analysis (PCA)during classification of leaf diseases.

Smitha.Naikwadi, 2013 proposed a texture based method namely spatial gray level based dependence matrix. However, the method has limitations of using advanced imaging characteristics for extraction of color features.

PiyushChaudry, 2012 proposed a color transformation based approach for detecting disease spots on the plant leaf. A median filter for smoothing combined with an OTSU based thresholding method was implemented in this work. Quantifying geometrical characteristics of these spots could have improved the overall performance of the technique.

Further, with the emergence of genetic algorithms to evolve programs to perform certain task was proposed by John Koza in 1992 (known as genetic programming). This method is mainly used in applications pertaining to optimization problems and regression problems. A trial and error method was used for effective MGGP based implementation.

The usage of SVM and SVR as an Artificial Intelligence (AI) method was further emphasized by Vijayaragahavan, 2017 in solving classification and regression problems respectively. The latter being extensively used for generalization ability to a solution model.

However, limited study is performed in classifying between leaf types using texture based imaging features which forms the basis for this study.

III. PROPOSED METHODOLOGY

The objective of the proposed method is to identify textural features to derive imaging characteristics that could discriminate between different leaf types particularly between Alternata-Alternata and Anthracnose.

The proposed methodology involves three stages mainly image preprocessing, image segmentation and classification (fig. 1). The dataset used in this study (Manu's Disease Dataset) contains 22 images of alteranata-alteranata and 23 images of anthracnose leaf type of which 20 images of alternata-alteranta and 20 images of anthracnose were considered which were observed to have varying patterns of disease structure, distribution and texture type.

Preprocessing:

The input image is separated into Red, Green and Blue component for which histogram equalization is performed for each channel for enhancing the contrast and is subsequently concatenated to obtain the RGB color model of the image. Further, the image is transformed from RGB model to Lab color space. A median filter is applied for each of the color component in order to perform image smoothing.

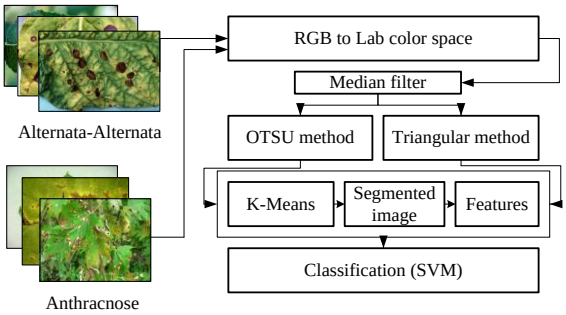


Fig 1: Proposed methodology

Segmentation using GLT with OTSU method:

Each channel component after smoothing is thresholded to gray level thresholding using OSTU technique followed by image binarization.The resulting binary image is further multiplied with the original RGB image to extract the intensity components of regions of interest. The imaging characteristics were derived from these regions which are further explained in the following section.

Segmentation using triangular method:

A histogram for gray level images is obtained that is used in thresholding factor in the triangular method proposed by Zach G.W, 1997. This method is effective when weak peaks are found from object pixels when viewed as histogram.A line is constructed from the lowest/highest value (based on context) (a) to the maximum of the histogram (b), A distance L which lies normal to the line and between the line and histogram is computed from a to b involving all values. The threshold value is defined as the distance between histogram and the maximal line. This method is based on a 1 dimensional histogram for threshold computation.

Region specific Imaging attributes:

The regions are converted from RGB to Lab color space for which Euclidean distance is measured for each channel component using k-means based clustering technique to obtain cluster centroids for variations in intensity values. The images are then segmented with the centroid obtained from k-means clustering, which is then converted from RGB color space to grayscale followed by binarization using the gray level thresholding. The following imaging characteristics pertaining to gray level co-occurrence matrix are derived from the segmented image:

a. Contrast: The contrast of image is defined as the gray scale color map used to enhance contrast of an image. This resulting map is an Mx3 matrix which is roughly distributed of gray scale histogram. If M is omitted, the current color map is used. An optimal result is obtained during image colors ordered by intensity. Contrast is computed as shown in eq. 1

$$Contrast = \sum_{i,j} |i-j|^2 p(i,j) \quad \ldots. (1)$$

$where, i, j \rightarrow image\ pixels$

b. Correlation: The correlation measure is defined as the measure of correlation strength between adjacent pixels of the segmented image. This is often used to measure the coherent patterns in image intensity. Correlation between pixels is computed as shown in eq. 2

$$Correlation = \sum_{i,j} \frac{(i-\mu i)(j-\mu j)p(i,j)}{\sigma_i \sigma_j} \quad \ldots. (2)$$

c. Energy: The energy is measured for indicating strength of pixel intensity through performing summation of squared elements in the GLCM. Energy is calculated as shown in eq. 3

$$Energy = \sum_{i,j} p(i,j)^2 \ldots. (3)$$

d. Homogeneity: The homogeneity measure is used to measure the proximity of the how the elements are distributed. In the GLCM. Homogeneity is computed as shown in eq. 4

$$homogeneity = \sum_{i,j} \frac{p(i,j)}{1+|i-j|} \quad \ldots. (4)$$

e. Mean and std. deviation: The mean of the segmented image is calculated as a measure of sum of all pixels over number of pixels in the whole segmented image. i.e.

$$\mu = \frac{sum\ of\ all\ pixels\ in\ image}{number\ of\ pixels\ in\ segmented\ image}$$

Where, $\mu \rightarrow$ mean

Further, standard deviation is measured across pixels over whole segmented image.

f. Entropy: A texture based feature namely entropy is measured in this experiment. It is basically a statistical method that measures the randomness in pixel intensity to characterize the textural attributes of the image.

g. Kurtosis: The measure produces a vector input, k which is the fourth central moment of the segmented image over the power of its standard deviation.

h. Skewness: This measure returns the skewness observed in the values of segmented image, which is defined as the third central moment of X that is divided by the cube of its standard deviation.

Other measures involve root mean square (RMS), variance and smoothness.

Classification using Machine learning:

An SVM based classification technique is used which is validated using a 5 fold cross validation technique. The SVM method uses a Gaussian kernel. Prior to classification, a PCA is applied to address the issue of dimensionality with explained variance of upto 95%.

The dataset considered in this experiment consists of a total of 40 samples with 20 samples of alternata-alternata and 20 samples with anthracnose. The data is divided into training and testing samples with 5 iterations, in which each iterationconsists of 32 training samples and 8 test samples. The hyper-parameters namely cost and gamma are varied to obtain an optimal global maxima for the training set that is further used in prediction of the test data.

IV. RESULTS AND DISCUSSION

The proposed methodology was able to classify the disease type between the two groups with a reasonably improved classification performance.

Table 1: Classifier performance

Classifier model	Accuracy	CP	AUC
SVM-OTSU	52.50	21/40	0.50
SVM-Triangular	70.0	28/40	0.63

As observed from table 1, The SVM based on triangular method of segmentation has an improved performance as compared to SVM-based OTSU method. A total of 28 of 40 samples were correctly classified in SVM-Triangular method as compared to SVM-OTSU method which only classified 21 of 40 samples correctly. This is further indicated in the receiver operating characteristic curve (Fig. 2) with Area under the Curve (AUC) of 0.63 as compared to 0.50 respectively.

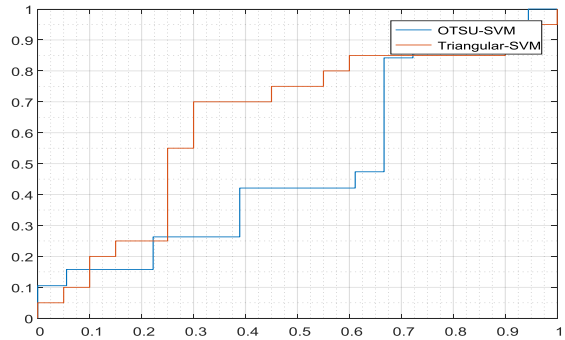


Fig 2: Receiver Operating Characteristic (ROC)

Fig 3 and Fig 4 indicates the area of affected regions in the leaf samples of the two groups. A comparative analysis is performed between the methods of segmentation. As indicated in Fig 3. the triangular method has an improved performance in identifying the affected region in group with an overall mean of 27.71 and 20.84 respectively Alternata-Alternata as compared to OTSU method. However, both the segmentation method indicates similar performance with OTSU having slightly improved performance in the case of Anthracnose group with an overall mean of 25.38 and 29.41 respectively.

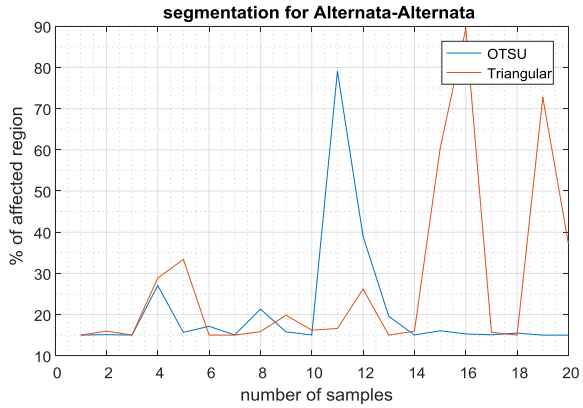


Fig 3: Segmentation for Alternata-Alternata

The difference between these variations in segmentation method indicates the improvement in the classifier performance.

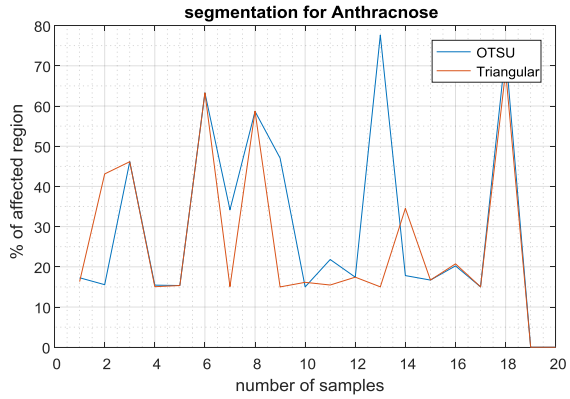


Fig 4: Segmentation for Anthracnose

As shown in fig 5, one sample image from alternata-alternata and one from anthracnose are considered. The image is originally in RGB color space which is then converted into Lab color space, The affected regions are indicated as dark intensity values. Upon further performing segmentation using OTSU method and binarization, the affected regions are shown in brighter pixel intensity as compared to non-affected region. Similarly, the affected regions are indicated with darker pixels in triangular method of segmentation.

As compared to other methods of segmentation, usage of machine learning techniques has improved the performance of segmentation of disease patterns with better generalization. Though this study has a comparatively lower classifier performance, the imaging characteristics of the fungi type across the leaves has been looked at a quantitative viewpoint.

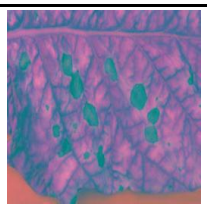
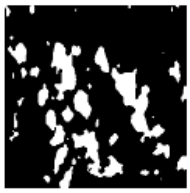
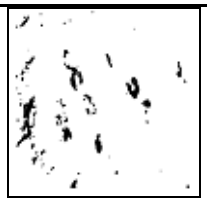
V. CONCLUSION

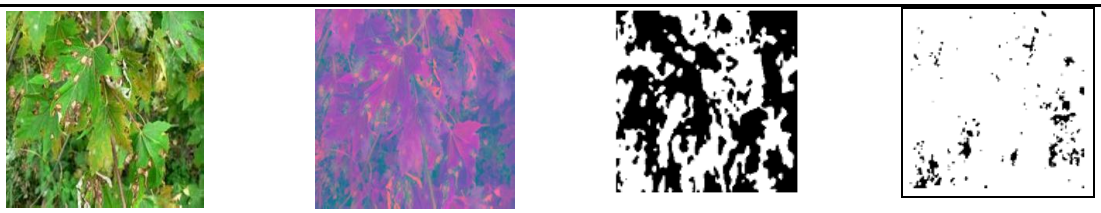
From the proposed study, patterns concerning the type of fungi such as alternata-alternata and anthracnose have been identified in leaves using a combination of segmentation technique with a classifier model with textural features pertaining to GLCM. The dataset considered has a total of 40 samples with 13 features.

The triangular based segmentation method showed a comparatively higher classifier performance as compared to OTSU based segmentation method. The features were extracted using a combination of k-means with RGB and Lab color space models along with an SVM based classification.

The proposed method showed an improved classifier performance in distinguishing Alternata-Alternata from Anthracnose with an overall accuracy of 70% using a triangular based segmentation with an SVM classifier, as compared to OTSU based segmentation with SVM. The proposed method correctly predicted 28 of 40 samples resulting in an AUC of 0.63.

Further research scope is in the direction of application of machine learning techniques in the field of agriculture. Of the many possible potential applications of machine learning techniques that could be used for the same, identifying and classifying types of leaf disease would result in improved prevention and better monitoring of plant diseases.

Input image	RGB2Lab	OTSU	Triangular
			
Alternata-Alternata			



Anthracnose

Fig 5: Resulting image for alternata-alternata and anthracnose segmentation

REFERENCES

1. Gavhale KR, Gawande U. An overview of the research on plant leaves disease detection using image processing techniques. IOSR Journal of Computer Engineering (IOSR-JCE). 2014 Jan;16(1):10-6.
2. Rumpf T, Mahlein AK, Steiner U, Oerke EC, Dehne HW, Plümer L. Early detection and classification of plant diseases with support vector machines based on hyperspectral reflectance. Computers and electronics in agriculture. 2010 Oct 1; 74 (1): 91-9.
3. Al-Hiary H, Bani-Ahmad S, Reyalat M, Braik M, ALRahamneh Z. Fast and accurate detection and classification of plant diseases. International Journal of Computer Applications. 2011 Mar 31;17(1):31-8.
4. Gulhane VA, Gurjar AA. Detection of diseases on cotton leaves and its possible diagnosis.International Journal of Image Processing (IJIP). 2011 Jun;5(5):590-8.
5. Patil JK, Kumar R. Advances in image processing for detection of plant diseases. Journal of Advanced Bioinformatics Applications and Research. 2011 Jun;2(2):135-41.
6. Kulkarni AH, Patil A. Applying image processing technique to detect plant diseases. International Journal of Modern Engineering Research. 2012 Sep;2(5):3661-4.
7. Revathi P, Hemalatha M. Classification of cotton leaf spot diseases using image processing edge detection techniques. In2012 International Conference on Emerging Trends in Science, Engineering and Technology (INCOSET) 2012 Dec 13 (pp. 169-173).IEEE.
8. Mokhled S. Al-Tarawneh. An Empirical Investigation of Olive Leave Spot Disease Using Auto-Cropping Segmentation and Fuzzy CMeans Classification.World Applied Sciences Journal. 2013;23(9):1207-11.
9. Kanjalkar HP, Lokhande SS. Feature extraction of leaf diseases. Int. J. Adv. Res. Comput. Eng. Technol.(IJARCET). 2014 Jan;3(1):153-5.
10. Chakravarti R, Meng X. A study of color histogram based image retrieval. In2009 Sixth International Conference on Information Technology: New Generations 2009 Apr 27 (pp. 1323-1328). IEEE.
11. Rasli RM, Muda TZ, Yusof Y, Bakar JA. Comparative analysis of content based image retrieval techniques using color histogram: a case study of GLCM and K-means clustering. In2012 Third International Conference on Intelligent Systems Modelling and Simulation 2012 Feb 8 (pp. 283-286).IEEE.
12. Redi M, Merialdo B. Saliency-aware color moments features for image categorization and retrieval. In2011 9th International Workshop on Content-Based Multimedia Indexing (CBMI) 2011 Jun 13 (pp. 199-204).IEEE.
13. Weng T, Yuan Y, Shen L, Zhao Y. Clothing image retrieval using color moment. In Proceedings of 2013 3rd International Conference on Computer Science and Network Technology 2013 Oct 12 (pp. 1016-1020). IEEE.
14. Doshi NP, Schaefer G. Texture classification using multi-dimensional LBP variance. In2013 2nd IAPR Asian Conference on Pattern Recognition 2013 Nov 5 (pp. 672-676). IEEE.
15. Belloulata K, Belallouche L, Belalia A, Kpalma K. Region based image retrieval using shape-adaptive dct. In2014 IEEE China Summit & International Conference on Signal and Information Processing (ChinaSIP) 2014 Jul 9 (pp. 470-474).IEEE.
16. [Bandaru R, Naik D. Retrieve the similar matching images using reduced SIFT with CED algorithm. In2014 International Conference on Control, Instrumentation, Communication and Computational Technologies (ICCICCT) 2014 Jul 10 (pp. 1242-1247). IEEE.
17. Chakravarti R, Meng X. A study of color histogram based image retrieval. In2009 Sixth International Conference on Information Technology: New Generations 2009 Apr 27 (pp. 1323-1328). IEEE.
18. Ma Z, Tavares JM, Jorge RN. A review on the current segmentation algorithms for medical images.InProceedings of the 1st International Conference on Imaging Theory and Applications (IMAGAPP) 2009.
19. Oliveira FP, Tavares JM. Matching contours in images through the use of curvature, distance to centroid and global optimization with order-preserving constraint.
20. Dhaygude SB, Kumbhar NP. Agricultural plant leaf disease detection using image processing.International Journal of Advanced Research in Electrical, Electronics and Instrumentation Engineering. 2013 Jan 1;2(1):599-602.
21. Badnakhe MR, Deshmukh PR. Infected leaf analysis and comparison by Otsu threshold and k-means clustering. International Journal of Advanced Research in Computer Science and Software Engineering. 2012 Mar;2(3).
22. Arivazhagan S, Shebiah RN, Ananthi S, Varthini SV. Detection of unhealthy region of plant leaves and classification of plant leaf diseases using texture features. Agricultural Engineering International: CIGR Journal. 2013 Apr 9;15(1):211-7.
23. Kulkarni AH, Patil A. Applying image processing technique to detect plant diseases. International Journal of Modern Engineering Research. 2012 Sep;2(5):3661-4.
24. Bashir S, Sharma N. Remote area plant disease detection using image processing. IOSR Journal of Electronics and Communication Engineering. 2012 Sep;2(6):31-4.
25. Naikwadi S, Amoda N. Advances in image processing for detection of plant diseases. International journal of application or innovation in engineering & management (IJAIEM). 2013 Nov;2(11).
26. Chaudhary P, Chaudhari AK, Cheeran AN, Godara S. Color transform based approach for disease spot detection on plant leaf. International Journal of Computer Science and Telecommunications. 2012 Jun;3(6):65-70.
27. Koza JR. Genetic programming.Vijayaraghavan P, Vosoughi S, Roy D. Twitter demographic classification using deep multi-modal multi-task learning. InProceedings of the 55th Annual Meeting of the Association for Computational Linguistics (Volume 2: Short Papers) 2017 Jul (pp. 478-483).
28. Zack GW, Rogers WE, Latt SA.Automatic measurement of sister chromatid exchange frequency.Journal of Histochemistry&Cytochemistry. 1977 Jul;25(7):741-53.

AUTHOR PROFILE

Mr. Shravankumar Arjunagi Received the Bachelor's degree in Electonics and communication from Visvesvaraya Technological University Belagavi, Karnataka 2002, M. Tech. degree in Computer Science and Engineering from VTU Belagavi, Karnataka in 2011. From August 2011 to till date working as an Assistant professor in Department of Computer Science and Engineering at APPA Institute of Engineering and Technology Kalaburgi- Karnataka. His research interests areas are, Image processing, Cloud Computing, IOT ,Data mining &presently pursing Ph.D under VTU, Belagavi, from 2015.

Dr. Nagaraj B. Patil Received the B.E. degree from The University of Gulbarga Karnataka 1993, M. Tech. degree from the AAIDU University of Allahabad in 2005,and the Ph.D. degree from the University of Singhania Rajastan India in 2012.He is having teaching experience of 27 years including 7 years research, from 2010 to 2018 .He worked as a Associate professor and HOD Dept. of CSE & ISE at Government College of Engineering, Raichur Karnataka. Presently he is working as a Principal of Government Engineering College Gangavati Karnataka. His research interests areas are Image Processing and Wireless sensor network.

.