

Detection of Breast Cancer using Digital Image Processing Techniques

E.Kodhai, S.Jaseema Yasmin, K.Subhasree, V.Vikneshwari

Abstract: Breast cancer, a major public health contention and currently causes an increased rate of cancer death in women. The preeminent intent of this project in medical diagnostics is by using mammography, that is a unique imaging technique in medicine for examining the breasts. A higher quality mammographic images (for example electronic pictures) are stored, mammography method (i.e exam) is performed, which is a prior stage for detection and diagnosis of bosom's malignant growth. In order to detect the tumor FCM algorithm for segmentation is used and the features are extracted by using multi-level wavelet transformation technique with PCA and then some features are added with GLCM features. Further, those segmented region features are extracted and the dataset is trained and tested completely. The images are classified by using SVM, KNN, Tree classifier, Neural Networks or Naive Bayes classifier. Finally, the images from Kaggle dataset are compared and categorized as normal, benign or malignant tumors.

Keywords: FCM; GLCM- Gray Level Co-Occurrence Matrix; Bosom; Classifier; Mammogram.

I. INTRODUCTION

Breast cancer that forms in the cells of the breasts that develops from bosom tissue, known as Cancer where Malignancy is a term that is used for a class of diseases to find or identify with great accuracy or precision by a typical development of cells and perforates the healthy cells. Big data is commonly defined as profoundly and astronomically immense data sets that have undergone growth beyond the competency to maintain and inspect them with conventional data processing tools and techniques. Big data defines a circumstance in which the data sets have grown to such gargantuan size which standard and statistics information disciplines can no longer efficaciously manipulate either the area of the position or the measurement and increase of the data set [1].

Big data is filled with Opportunities and Occurrences. As Stated by IBM, on a daily basis we people engender 2.5 quintillions of data every day i.e.; 2.5×10^{18} bytes, this is calculated as huge as 90 percent among the data of the world compared with last two years, which are emanated from everywhere including sensors that are used to sense the climatic changes, Posting digital images and videos on

social media sites and stores information on online transaction, that are recorded and retrieved in online purchases, and mobile phone using Global Positioning System signals. The Agent used for Big data, with many indicated fact that all those data have essential value that can be used to extend the application using analytics, algorithm, and other methods. National Corporations such as the National Oceanic and Atmospheric Administration (NOAA), The National Aeronautics and Space Administration (NASA), multifarious medicinal drug organizations, various power companies have an incapacitated process or fluctuation of the enormous amounts of data and now grip Big data skills on a regular support to fragment value from them. In today's business, Big data plays a major part in Immensely colossal institutions effectively overlook the desideratum to support highly immense set of organized and unorganized data-from data warehouses for undertaking facts regarding the tweets of employee, from the records of the supplier to regulatory filings—to comply with the rules and regulation of regime, that need has been changed and considered by the current justice victims that have enheartened business agencies to continue with sizably voluminous amount of reports, electronic mails, and other transmissions that are made digital, such as multimedia messaging services and telephony devices, that might be in need for E- Revelation if it is being overlooked on legal action [1].

Big data analytics have improved healthcare by giving customized drug and prescriptive investigation. Scientists are mining the data to recognize what medicine is progressively attentive for certain conditions, recognize design indicated with prescribed responses, and increases other vital data that can be used to encourage sufferers and decreases prices. With the additional selection of M-health, E-health, and acceptable innovations, the volume of information is expanding at an enormous value. This incorporates computerized wellbeing to store data, imaging and processing data, victim engender data, sensor data, and other configurations of data [2].

Wavelets are always being given in its current form, where the input does not have any wavelet transforms. A set of wavelets-cognate function compatible with the wavelets toolbox is developed which will sanction the utilizer to implement all the wavelet-transform concepts. Image segmentation is the set of input functions available for image segmentation are performed with the fundamental and elongated transformation functions. Some of the incipient functions for Hough transform processing and region growing are additionally developed.

Revised Manuscript Received on June 22, 2019.

E.Kodhai, Department of Computer Science and Engineering, Sri ManakulaVinayagar Engineering College Puducherry, India

S.Jaseema Yasmin, Department of Computer Science and Engineering, Sri ManakulaVinayagar Engineering College Puducherry, India

K.Subhasree, Department of Computer Science and Engineering, Sri ManakulaVinayagar Engineering College Puducherry, India

V.Vikneshwari, Department of Computer Science and Engineering, Sri ManakulaVinayagar Engineering College Puducherry, India

Image processing sometimes works as a difference or contrast between similar things that are being defined with a disciplined concept in which both input based images and output images are processed. The field of Artificial Intelligence beginning from its earlier stages and started developing gradually than being slowly anticipated. Image analyzing processing (adscitiously called as image understanding) is performed with different computer vision and image, where there are three types of computerized process in this duration: low-, middle-, and high- level processes. Low-level processes involve some of the preliminary operations such as image sharpening and polishing, image, and, enhancing and controlling contrast, Image processing in decreasing noise. Mid-level process on images involves implementing through Pre-Processing, Segmenting all of which make partitions of an image into regions or objects, and provides techniques for reducing those regions into a type of individual components for system processing and system classification in recognition of those objects.

Conclusively, the higher- level processing involves a method of applying sense through intelligence agents to have a collection of the recognized objects, as in analyzing images, and, at the slow changeover of time, in performing the intellectual functions mundanely associated with the vision of human[8].

MATLAB is a high-level performance Software for handling technical computations which ensembles along with computation and coding can be done easily along user-friendly environment, where solutions to the problems are expressed in some familiar mathematical notations. MATLAB is a name that stands for the matrix laboratory to provide an easy way of use and access to perform matrix using the development by the LIN pack (Linear System Package) and EIS pack (EIGN system package) project. Today, MATLAB engines incorporate the LA pack (Linear Algebra Package) and BLAS (Basic Linear Algebra Subprograms) libraries, creating software for computing matrices[2].

II. RELATED WORKS

Siyabend Turgut et al., [3] The classification of the patients for the Microarray breast cancer data has different methods selection methods such as Recursive Feature Elimination (RFE) And Randomized Logistic Regression (RLR) are being implemented and it performs classifications are compared with the output of the first cases. There are different layers of neurons that analysis the changes of the layers and the classification on the accuracy.

Varalatchoumy.M et al., [4] Histogram Equalization and Techniques such as Dynamic Thresholding for Pre-processing. A Region of interest (ROI) was separated after performing the preprocess with k-means clustering and Particle Swarm Optimization methods. Also Preprocessing was done using different morphological functions such as Erosion and Dilation. For the identification of ROI, a changed situation of segmentation and also another method for preprocessing using histogram equalization and an advanced level set approach for performing segmentation with different morphological operations and contrast limited

adaptive histogram equalization for Image Preprocessing are used.

Ammu P K et al., [6] This paper reviews the uses of some common and few major feature Selection techniques that are employed in the micro array data from DNA which is one of the most important procedures in bioinformatics. Biogeography-Based Optimization (BBO) and using Particle Swarm Optimization (PSO). Redundancy Based Feature selection approaches and a two-stage hybrid filter wrapper Method is used for Geneselection based on the dependency of features where the features are classified as independent, half dependent and dependent features.

Bing Lan et al., [7] A novice fuzzy stage set algorithm is suggested in this paper to hastenunforced medical scan segmentation. It is capable of directly expand from the starting segmentation by Spatial Fuzzy Clustering. The managing parameters of level set advancement are also approximated from the outcomes of fuzzy clustering. The stage set procedures employ dynamic evolutionary boundaries for image segmentation. It utilizes a fuzzy-c means (FCM) with spatial limitations is used to establish the estimated contours in a medical image. The results confirm its efficacy for medical image segmentation.

Anna N. Karahaliou et al., [8] This paper investigates some texture featuresof the tissue which is covered the micro calcification (MC) clusters on Mammograms for breast cancer analysis. The tissue that surrounds MCS is mentioned as that on the real and wavelet decomposed pictures, depend on a redundant discrete wavelet transform. Gray-level texture and wavelet coefficient texture characteristics at 3 decomposition stages are drawn from surrounding tissue regions of interest (ST-ROIS). It is under the receiver operating characteristic curve (AZ) of 0.989. Outcome suggest that its texture diagnosis can provide to the diagnosis of breast cancer aided by systems.

Sangeetha R et al., [2] The feature extraction process in this system is done by using GLCM (Gray Level Co-Occurrence Matrix) and by the initial order statistics features. To divide the mammographs as abnormal or normal, two divided operations are classified by using KNN algorithm (k-nearest neighbor) to detect the early stage of the tumor and also to find the accuracy of tumor region. The image dataset is processed and generated from the MIAS (Mammographic Image Analysis Society). The experimental results in this method give an accuracy of 92% which can be further improved and the segmentation method includes the basic otsu's thresholding algorithm. Due to the lack of detection algorithms, the computational time used is higher in this early level detection system for breast cancer.

III. PROPOSEDSYSTEM

The Proposed system is created using MATLAB software which is used as an friendly interface for users. The important steps used in this system are to input the data, perform preprocessing, image segmentation, extraction of features, dataset creation, and clustering. kaggle dataset is used for obtaining dataset and experimenting results.

This system is useful in finding tumour region or detecting cancer tissues in the mammography as shown in Fig.1 which is the block diagram for the proposed system that is designed in the working of breast cancer detection.

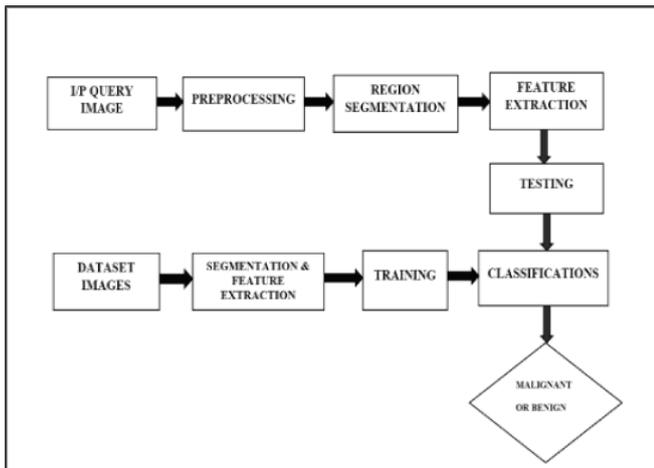


Fig. 1 Block diagram of detection of breast cancer

Preprocessing

An input image is Preprocessed which is a stage that the digital image is checked with effulgence characteristics of a pixel with an impact of background noise which ameliorates the quality of the image and make it suitable for various processing without the any disturbance of noise or any unnecessary components in the background of the mammogram by image processing. The important part of pre-processing is the improvement of the image data that undergoes undesirable distortions or in the enhancement of some image highpoints for performing further processes. The Final segmentation of the image is followed by pre-processing of the image that involves steps to calculate the size and type of the tumor that may involve enormous optimization practices to further ameliorate the results.

Segmentation

The goal of segmentation is to simplify and transmute the illustration of an image to clearly analyze. Its concept is to notice the position of the doubtful region to avail and to relegate the malformations classified into cancerous or non-cancerous tumours. FCM (i.e,Fuzzy-C means clustering) is utilized in segmenting the tumour affected region by fixing the number of clusters which begins in the estimation for the center parts of the cluster, that are envisioned to mark the mean location of each cluster. The functional consequences of the pixel represents the chances that it belongs to a concrete cluster. The pixels of high membership values are proximated to the centroid of their clusters, and low membership values are made away from the centroids that are observed as pixels with the data. Convergence for the membership function can be checked by distinguishing the transmutations in the cluster center at two successive iteration steps.

Feature extraction

Feature extraction is performed on dissevering a dataset into a training set and testing set where the large number of data is processed for training, and a more minute chunk of the data is utilized for testing. Analysis accommodations

arbitrarily sample the data to ascertain that the testing and training sets are homogeneous. The Multi-Level Wavelet Transformation method with the PCA (Principle Component Analysis) is a quantifiable technique that utilizes some of the orthogonal transformations to transmute over an abundant possible observations that are made as associated variables(entities each of which goes up against different numerical values) that are to be linearly uncorrelated as different variables are called the principal components which are used to extract the features of the Segmented Region. The GLCM Features with some additional features are also to be extracted the fig.2 shows the features and its density. After the process of extracting features, Testing of the dataset is performed as the process of executing a program or application with the intent of identifying bugs in working or requirement analysis.

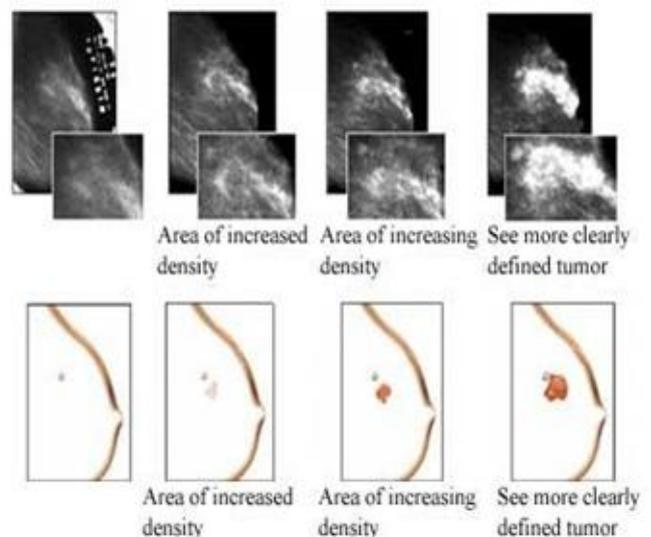


Fig. 2 Extracted Features

Testing and training

The dataset is completely trained and with the use of analytical data engine which apparently separates small part or quantity intended to show what the whole is like for the data to assure that the testing and training sets are comparable, which are later classified using the different classification methods such as K-Nearest Neighbour algorithm (knn) is used for classification and regression notified as non-parametric technique or SVM (Support-Vector Machine) that are predicated on defining the decision boundaries by the conception of decision planes. A plane that is chosen for the set of objects to isolate having different class involvement or using tree classifier is a tree learning that utilizes a decision tree to go from observation to decide the item's goal value. Statistics utilized in data mining and machine learning or otherwise by using neural networks or Bayes classifier are the predictive modeling approaches that is used for separating data and comparing the query image with the given dataset.

The result established here is the type of tumour state of the breast which may be given as normal, benign (i.e; extra growth of tissues) or malignant, to be stated as the output of consideration.



Fig 3 represents a rough MATLAB implementation program.

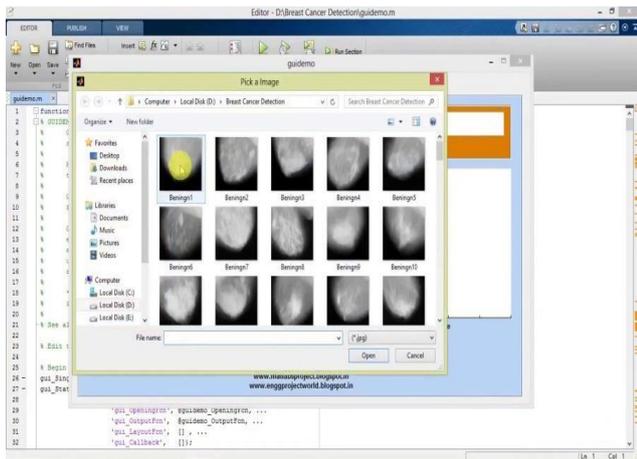


Fig. 3 Implementation Process

The proposed system is preceding from the given detection system in different influencing factors such as; the raw attributes are compared as compound valued replications of a set of Multi-Resolution Gabor Filters additionally the cross operator provides a simple approximation to the gradient magnitude and classical operator is detecting Edges and their Orientations Methods on improving ACCURACY and COMPUTATIONAL TIME. It stabilizes the utilization of caliber set evolution that spontaneously proceeds the particular boundaries when it is approached once, which prevents the dissemination of any boundary leakage.

IV. EXPECTED OUTPUT

The input query image to the system is being pre-processed and it is converted into binary form, Region segmentation is performed using FCM algorithm - which commences with an initial prediction in grouping a set of objects, which are determined to note the mean location of each cluster and the use of this FCM allocates every data point with a particular grade for each and every cluster.

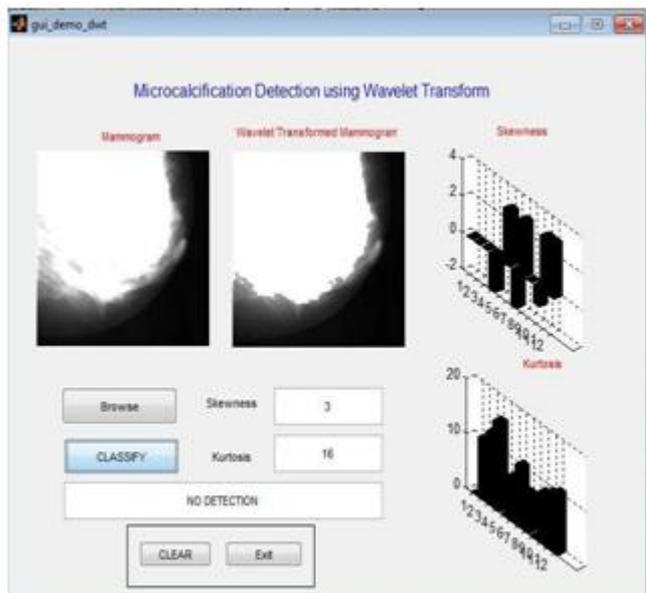


Fig. 4 Detection using Transformation

Feature extraction is done through multi-level wavelet transformation with PCA and the additional features with the GLCM are added to it. After that features are extracted & process dataset is completely trained and classified by using the either by KNN or By Tree Classifier. Svm, Neural Network or other machine learning concept. Figure 4 and Figure 5 describes the tumour detection using transformation for the mammography. The breast cancer or the tumour region is detected which the tumour is a mass of abnormal tissue.

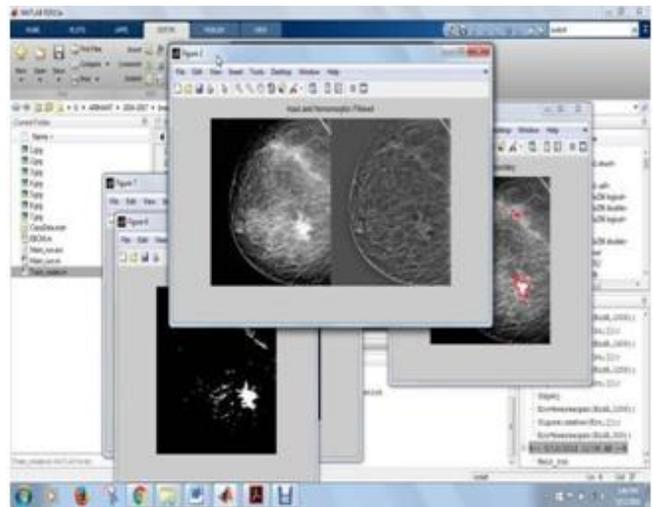


Fig. 5 Implementing Mammography

There are many types of breast cancer and the most mundane types among them are ductal carcinoma in situ, invasive ductal carcinoma, and invasive lobular carcinoma which includes two types of breast cancer tumours: those that are non-cancerous, or 'benign', and those that are cancerous, which are 'malignant' which can be processed and detected as Shown in the Figure.6 which can be later used to predict the cancer and can be found with the affected stage and provided with a curative treatment.

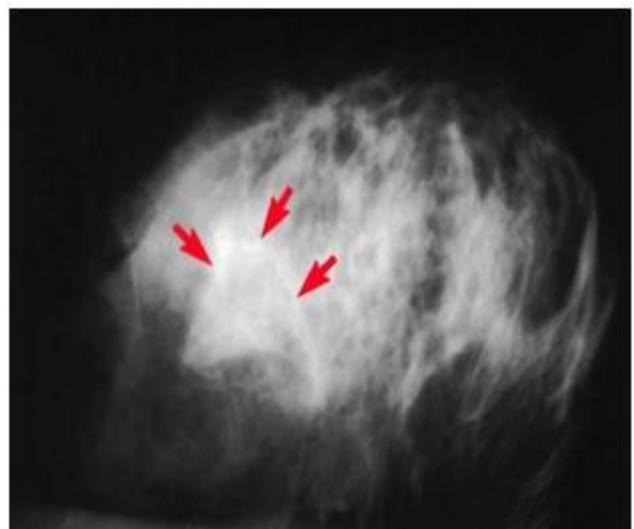


Fig. 6 Detection of Tumour Region in Breast

V. CONCLUSION

The research works that are present now are only used for detecting the breast cancer with swelling in body parts, Generally without inflammation, caused by an abnormal growth of tissue, whether benign or malignant stage and are not with a quality or state of being correct or precise to 100% which are highly dangerous. By adding furthermore attributes that cause breast cancer is added to the previously used influencing attributes and prediction of breast cancer can be performed using a suitable feature selection mechanism to obtain the result with higher accuracy. Researches that are made on breast cancer with the help of unique process called digital image detection. They are not newly innovated but lack of proper methods for early detection at micro calcification stage in medical domain is still a challenge. So, by using FCM (fuzzy-c-means) clustering for tumor region finding and segmentation is completely analyzed by using the Multi-Level Wavelet and PCA along with GLCM features based on it. Later the features are extracted and process dataset is completely trained and classified by using the either by KNN or SVM or tree classifier of machine learning concept to detect the type of tumour formed.

REFERENCES

1. Salama, G. I., Abdelhalim, M., &Zeid, M. A. (2012). Breast cancer diagnosis on three different datasets using multi-classifiers. International Journal of Computer and Information Technology, 1 (1), 0764–2277.
2. Sangeetha R , Dr. Srikanta Murthy K “A Novel Approach for Detection of Breast Cancer at an early stage using Digital Image Processing techniques” International Conference on Inventive Systems andControl(ICISC-2017).
3. Siyabend Turgut, Mustafa Dagtekin, TolgaEnsari“ Microarray Breast Cancer Data Classification Using Machine Learning Methods ”978-1-5386-5135-3/18/\$31.00 ©2018 Ieee.
4. Varalatchoumy.M, Ravishankar.M “Four Novel Approaches For Detection of Region of Interest In Mammograms” A Comparative Study Proceedings Of The International Conference On Intelligent Sustainable Systems (Iciss 2017) Ieee Xplore Compliant - Part Number: Cfp17m19-Art, Isbn:978-1- 5386-1959-9.
5. Ameer Nisha.S, ShajunNisha.S, Dr.M.MohamedSathik “A Study On Surf &Hog Descriptors For Alzheimer’s Disease Detection” (Ijret 2017) E-Issn: 2395 - 0056, P-Issn:2395-0072.
6. AmmuPK,PreejaV “Review On Feature Selection Techniques Of DNA Microarray Data” International Journal Of Computer Applications, Volume 61– No.12, January2013.
7. Libn, Chui Ck, Chang S, Ong Sh “Integrating Spatial Fuzzy Clustering with Levelset Methods For Automated Medical Image Segmentation” Elsevier Ltd. 0010- 4825/©2010.
8. Anna N. Karahaliou, Ioannis S. Boniatos, Spyros G. Skiadopoulos, Filippos N. Sakellaropoulos, Nikolaos S. Arikidis, Eleni A. Likaki, George S. Panayiotakis, Lena I. Costaridou “Breast Cancer Diagnosis: Analyzing Texture of Tissue Surrounding Microcalcifications” 1089-7771/© 2008Ieee.
9. P.M.Sampat, M.K.Markey, A.C.Bovik, “Computer- Aided Detection And Diagnosis In Mammography,” In Handbook Of Image and Video Processing, A. C. Bovik, Ed.,2nd Ed. New York: Academic, 2005, Pp. 1195–1217.
10. L .Tabar, P. Dean., “Mammography And Breast Cancer; The New Era,” International Journal Of Gynecology And Obstetrics, Vol.28, Issus 3, September 2003, Pp.319-326.
11. Gaurav Kumar And Pradeep Kumar Bhatia, “A Detailed Review Of Feature Extraction In Image Processing System” International Conference On Advanced Computing and Communication Technologies, 2014, Pp.5– 12.
12. A.Qayyum and A.Basit, “Automatic Breast Segmentation And Cancer Detection Via Svm In Mammograms”, Ieee,2016.
13. D.Dinsha, “Breast Tumor Segmentation And Classification Using SVM And Bayesian From Thermogram Images., ”Unique Journal Of Engineering & Advanced Sciences, Vol. 2, 2014, Pp:147-151.J.
14. K.N.Nyein Hlaing and Anilkumar K.G, “Myanmar Paper Currency Recognition Using GLCM And Kn”, Second Asian Conference On Defence Technology (Acdt), Ieee, 2016.
15. O.S.Pianyk, Digital Imaging And Communications In Medicine: A Practical Introduction And Survival Guide, 1st Ed. Springer Publishing Company, Incorporated,2008.
16. M. J. Weinberger, G. Seroussi, And G. Sapiro, “The Loco-I Lossless Image Compression Algorithm: Principles And Standardization Into Jpegs” Ieee Transactions On Image Processing, Vol. 9, No. 8, Pp. 1309–1324, Aug2000.
17. Iso/Iec 14495-1, Itu-T Rec. T.87, Information Technology – Lossless and Near-Lossless Compression Of Continuous-Tone Still Images: Baseline,1998.
18. Manikanthan, S.V., Padmapriya, T. A secured multi-level key management technique for intensified wireless sensor network. International Journal of Recent Technology and Engineering. 2019