Educational Administration: Theory and Practice
2024,30(4), 2989-2995
ISSN:2148-2403
https://kuey.net/

Research Article

Analysis Of Feature Extraction Techniques In Different Contrast Enhanced CT Images Of Liver

Rituparna Sarma[1], Yogesh Kumar Gupta[2]

[1]PhD Scholar,[2]Assistant Professor Email:- rituparna.sarma96@gmail.com, Email:-gyogesh@banasthali.in
[1,2]Department of Computer Science and Engineering Banasthali Vidyapith

Citation: Rituparna Sarma (2024), Analysis Of Feature Extraction Techniques In Different Contrast Enhanced CT Images Of Liver *Educational Administration: Theory And Practice*, *30*(4), 2989-2995
Doi: 10.53555/kuey.v30i4.1970

ARTICLE INFO

ABSTRACT

Feature extraction is an important part of segmentation techniques applied to images. Extraction of various features mean identification of different attributes that characterize an image. This process is quite challenging because of the image resolution and its complexity. In this paper we are trying to detect the cancerous tissues from the liver organ where the extraction of tissue features further requires differentiating between the cancerous and non-cancerous tissue patches. It is important to identity texture features that best describe a healthy and an unhealthy tissue from the digital image. Also, it is necessary to include a good number of texture features for better classification. In this paper, two feature extraction techniques, namely Gray-Level Co-Occurance Matrix (GLCM) and Gray-level run-length matrix (GLRLM) are used for identifying the texture characteristics of tumor in liver organ. These techniques depend on the spatial distribution of intensity values or gray levels in the liver region. The extracted features are then classified using SVM classifier. The accuracy of the model is satisfactory and effective for tumor diagnosis and decision making process for treatment of tumor.

Keywords: feature extraction, GLCM, GLRLM, CT image, Histopathological image analysis, ROI, Marker Controlled watershed segmentation

Introduction:

Medical Imaging is a technique which uses the ultrasound images of the different organs of the body to understand the abnormalities present in them. Computed tomography (CT) images are one such medical imaging technique to identify the existence of cancerous cells in an organ [1]. However, the USG images are of low quality as they are always accompanied by noise speckles [2]. Hence it is important enough to remove the unwanted noise from the CT image as a pre-processing step. There are many pre-processing techniques which can be used to remove the noise from the ultrasound image. This work uses median filter as a pre-processing step to filter the input medical image. The filtered image then goes through the segmentation process. Various segmentation techniques are available which help segment and identify the Region of Interest (ROI). The segmentation process used for our work is Marker Controlled watershed transform. It has been found in research that combination of marker controlled watershed transform and median filter improves the performance of the segmentation algorithm [3]. Once the segmentation of the medical image is completed, the feature extraction process starts. Feature extraction basically extracts the most important features based on their pixel intensity relationship.
Automation of Histopathological image analysis have gathered immense popularity in recent times due to the increase in the number of cancer cases [4]. The texture features of the medical images automate the classification process as well as enhances the decision-making process [5]. Thus, it helps in the diagnosis and treatment monitoring process of cancer patients.

Proposed Methodology:

The entire flow of proposed methodology starts with the acquisition of the CT images of patients. The patient CT scans are preprocessed to remove the unwanted noise and then passed through the segmentation phase. The segmented images are taken as input for extraction of necessary features using GLCM and GLRLM feature

extraction techniques. Thereafter, the classification of the images into two categories of tumored and untumored is the final outcome of the methodology used.

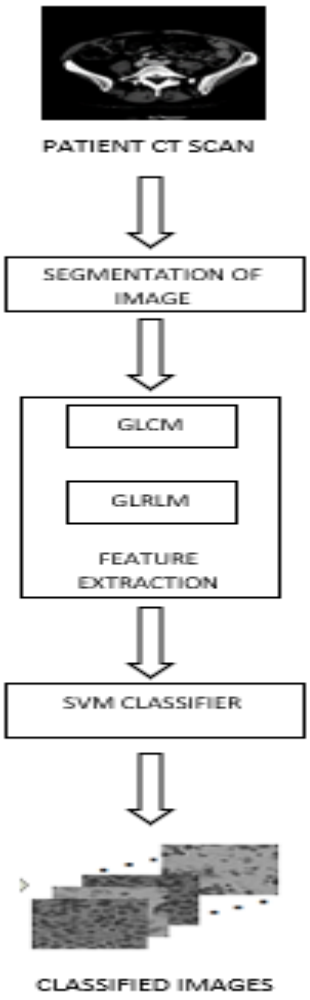


Figure1: System Overview

Feature Extraction Methods:

Feature extraction is a method of simplifying the task of representing a dataset with large number of resources and characteristics as it incurs lot of memory as well as computational power with complex datasets. It also overfits the classification model used for classifying the complex datasets. Feature extraction techniques helps finding the underlying features of a particular image and then tries to represent those features in a unique form, so that they can be used for robust, accurate classification and segmentation of objects. The statistical texture analysis is classified into first-order, second-order and higher-order statistics depending on the number of intensity points or pixels present in each combination. GLCM is a second order statistical feature extraction algorithm while GLRLM is a higher order statistical feature extraction model.

I. ***Gray level co-occurrence matrix (GLCM):*** GLCM or spatial gray level dependency matrix is a frequently used method for texture feature extraction for analysis in medical images. It calculates the frequency of occurrence of pixel pairs with specific values [7,8,9]. The output of this technique is a 2D matrix consisting of the same number of rows and columns as that of the gray values in the image. GLCM is very sensitive to the size of the image data in which the estimation is made. Some of the features extracted in this technique are described below:

a) Energy: It is also known as uniformity or angular second moment and it calculates image homogeneity. It calculates the sum of squared elements. Energy is high when pixels are very similar.

$$Energy = \sum_{i,j=0}^{levels-1} P_{ij}^2$$

b) Contrast: It measures the local variations in the GLCM.

$$Contrast = \sum_{i,j=0}^{levels-1} P_{i,j}\,(i-j)^2$$

c) Homogeneity : It measures the closeness of the distribution of elements in the GLCM to its diagonal.

$$Homogeneity = \sum_{i,j=0}^{levels-1} \frac{P_{i,j}}{1+(i-j)^2}$$

d) Correlation : It measures the joint probability occurrence of the specified pixel pairs.

$$Correlation = \sum_{i,j=0}^{levels-1} P_{i,j}\left((i-\mu_i)(j-\mu_j)\middle|\sqrt{\sigma_i^2\sigma_j^2}\right)$$

e) Inverse Difference : It is local homogeneity and calculated as follows

$$Inverse\ Difference = \sum_{i,j=0}^{levels-1} \frac{P_{i,j}}{1+(i-j)^2}$$

f) Entropy: It measures the part of information required to compress the input image.

$$Entropy = \sum_{i,j=0}^{levels-1} -P_{i,j} * \log P_{ij}$$

g) Dissimilarity: It is calculated as follows

$$Dissimilarity = \sum_{i,j=0}^{levels-1} P_{i,j}\,|i-j|$$

where *i, j* are the spatial coordinates of the function $P_{i,j}$ and *levels* is the number of gray levels in the input image.

μ is the GLCM mean and calculated as

$$\mu = \sum_{i,j=0}^{levels-1} iP_{i,j}$$

σ² is the variance of the intensities of all reference pixels in the relationships that contributed to the GLCM

$$\sigma^2 = \sum_{i,j=0}^{levels-1} P_{i,j}(i-\mu)^2$$

The features extracted using GLCM algorithm for this research are autocorrelation, cluster prominence, cluster shade, contrast, correlation, difference variance, difference entropy, dissimilarity, energy, entropy, homogeneity, information measure of correlation1, information measure of correlation2, inverse difference, maximum probability, sum average, sum entropy, sum of squares (variance) and sum variance [10].

II. *Gray Level Run Length Matrix (GLRLM):* Gray level run length matrix is a technique to extract the texture features of images to analyse their characteristics in order to use them for classification model [6,7]. The output of this technique is a 2D matrix where each element gives the total number of occurrences of the gray level in the given direction.

The features extracted using GLRLM algorithm[10] are calculated as follows considering $P_{i,j}$ is the input image matrix.

a) Short Run Emphasis (SRE):

$$SRE = \sum_{i=1}^{C}\sum_{j=1}^{R} \frac{P_{i,j}}{j^2}$$

b) Long Run Emphasis (LRE):

$$LRE = \sum_{i=1}^{C}\sum_{j=1}^{R} j^2 P_{i,j}$$

c) Gray level non-uniformity (GLN):

$$GLN = \sum_{i=1}^{C}\left(\sum_{j=1}^{R} P_{i,j}\right)^2$$

d) Run length non-uniformity (RLN):

$$RLN = \sum_{i=1}^{R}\left(\sum_{j=1}^{C} P_{i,j}\right)^2$$

e) Run Percentage (RP):

$$RP = \frac{1}{n}S$$

f) Low Gray Level Run Emphasis (LGLRE):

$$LGRE = \sum_{i=1}^{C}\sum_{j=1}^{R}\frac{P_{i,j}}{i^2}$$

g) High Gray Level Run Emphasis (HGLRE):

$$HGRE = \sum_{i=1}^{C}\sum_{j=1}^{R} i^2 P_{i,j}$$

SVM classifier:

The features extracted is then fed into a classification model to perform the learning task to categorize the data into its destined classes. SVM is a supervised learning technique which creates a decision boundary or hyperplane to classify the input data into its correct classes [11]. It has been widely used in the field of medical image segmentation as it categorizes the classes with higher accuracy than other traditional learning approaches[12].

The rest of the paper consists of the materials and methods used for the research and results found while performing the experiments. Finally, the paper is concluded with the analysis of feature extraction techniques and the SVM classifier accuracy applied for my research work.

Materials and methods:

The techniques used are part of image processing and software used are MATLAB R2019a and python programming. The dataset used for this research is publicly available liver and liver tumor clinical dataset so that real time data would be used to evaluate the proposed model. The proposed method is evaluated on publicly available dataset 3Dircadb from Research Institute against Digestive Cancer (Ircad 2016) tumors dataset. All datasets used in tumor segmentation are acquired at different enhancement phases with various scanners.

Experimental Results:

In this research, two different datasets have been used. Dataset1 consists of 129 CT images of patients having tumor in their liver organ while dataset2 consists of 139 CT images of patients having no tumor in their liver organ. The size of the CT images used for this research is 512 X 512 pixels. The voxel size of dataset1 is 0.57 x 0.57 x 1.6 mm. The voxel size of dataset2 is 0.78 x 0.78 x 1.6 mm. The input CT images taken from the medical organisations often has noisy elements. This can reduce the accuracy level of the segmentation and classification model. Hence in order to get rid of these disturbances, image is first filtered or preprocessed to get an image with high contrast and low or no noise. This improved image then goes through the segmentation process and then features are extracted using GLCM and GLRLM techniques for better classification of the entities.

The GLCM and GLRLM feature values of one set of tumored and untumored CT images have been presented in the table 1 and table 2 below.

	GLCM Features	**value**
Tumored liver	Autocorrelation	1.3186
	clusterProminence	1.0563
	clusterShade	0.5840
	Contrast	0.0079
	Correlation	0.9590
	differenceEntropy	0.0459
	differenceVariance	0.0078
	Dissimilarity	0.0079
	Energy	0.8003
	Entropy	0.3834
	Homogeneity	0.9961
	informationMeasureOfCorrelation1	-0.8765
	informationMeasureOfCorrelation2	0.6710
	inverseDifference	0.9961
	maximumProbability	0.8886

	sumAverage	2.2150
	sumEntropy	0.3779
	sumOfSquaresVariance	0.0959
	sumVariance	0.3759
Untumored Liver	Autocorrelation	1.3743
	clusterProminence	1.1506
	clusterShade	0.6438
	Contrast	0.0066
	Correlation	0.9701
	differenceEntropy	0.0396
	differenceVariance	0.0065
	Dissimilarity	0.0066
	Energy	0.7734
	Entropy	0.4153
	Homogeneity	0.9967
	informationMeasureOfCorrelation1	-0.9025
	informationMeasureOfCorrelation2	0.7035
	inverseDifference	0.9967
	maximumProbability	0.8709
	sumAverage	2.2517
	sumEntropy	0.4107
	sumOfSquaresVariance	0.1100
	sumVariance	0.4335

Table 1: GLCM Features

	GLRLM Features	**Value**
Tumored liver	SRE	0.0921
	LRE	2.7746e+04
	GLN	5.5509e+03
	RP	0.0401
	RLN	160.9190
	LGRE	98.5157
	HGRE	5.5509e+03
Untumored liver	SRE	0.1238
	LRE	3.2280e+04
	GLN	5.0712e+03
	RP	0.0362
	RLN	210.4757
	LGRE	95.2893
	HGRE	5.0712e+03

Table 2: GLRLM Features

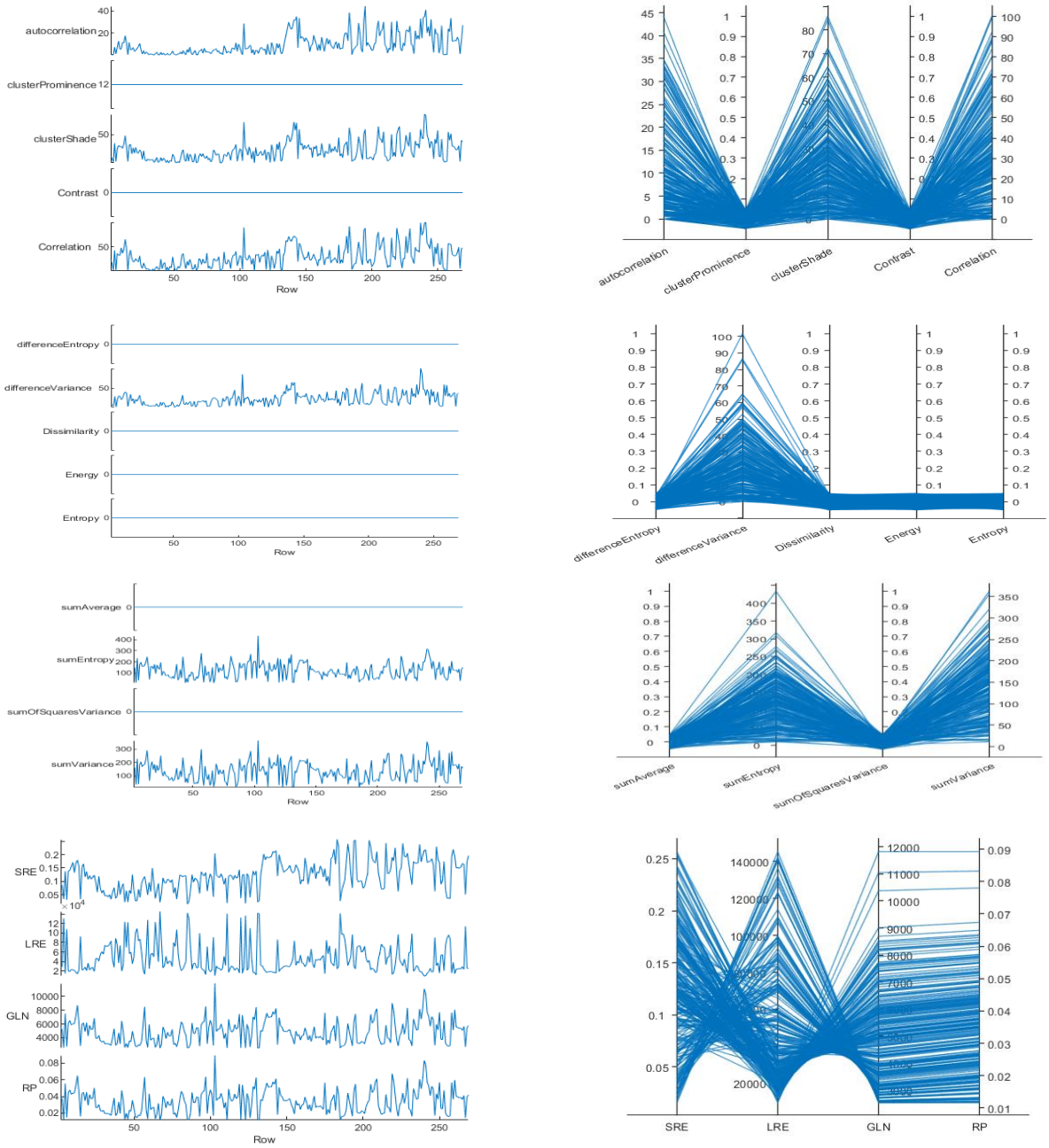
After feature extraction using GLCM and GLRLM, SVM classifier classifies the tumored and untumored liver CT images of patients with 86% accuracy in case of tumored tissues and 96% accuracy in case of untumored tissues respectively. Thus, the feature extraction from GLCM and GLRLM techniques are highly effective and accurate in order to classify the CT images of patients' tissues.

autocorre	clusterPro	clusterSha	Contrast	Correlatio	differenc	differenc	Dissimilar	Energy	Entropy	Homogen	informati	informati	inverseDi	maximum	sumAvera	sumEntro	sumOfSqu	sumVariat
0.544269	12	11.46421	0	23.28532	0	8.619432	0	0	0	112	0	0	0	0	0	158.8598	0	162.2946
0	12	8.200934	0	16.46186	0	5.912334	0	0	0	112	0	0	0	0	0	143.6364	0	158.9132
0	12	1.499947	0	1.688761	0	3.776334	0	0	0	112	0	0	0	0	0	11.15346	0	33.52274
6.954182	12	34.70723	0	42.35501	0	28.55866	0	0	0	112	0	0	0	0	0	232.5973	0	225.6993
0.909928	12	2.38481	0	0	0	0.915565	0	0	0	112	0	0	0	0	0	32.39552	0	37.04069
6.715369	12	24.71964	0	36.22064	0	20.12867	0	0	0	112	0	0	0	0	0	177.5798	0	183.3804
12.55288	12	30.65329	0	32.74737	0	30.12119	0	0	0	112	0	0	0	0	0	196.2535	0	204.4317
5.256737	12	29.01484	0	32.01617	0	28.20769	0	0	0	112	0	0	0	0	0	159.5321	0	167.0275
5.334035	12	28.8908	0	47.53784	0	19.91679	0	0	0	112	0	0	0	0	0	156.667	0	181.158
11.36775	12	25.81624	0	36.34898	0	24.47379	0	0	0	112	0	0	0	0	0	169.2785	0	182.8019
10.97352	12	33.48825	0	44.47175	0	34.8848	0	0	0	112	0	0	0	0	0	198.6471	0	221.7635
17.50018	12	48.06065	0	65.15862	0	40.00665	0	0	0	112	0	0	0	0	0	245.3956	0	289.1136
10.30724	12	27.86732	0	47.15266	0	37.23033	0	0	0	112	0	0	0	0	0	163.5226	0	193.743
0	12	2.455701	0	0.85156	0	1.643738	0	0	0	112	0	0	0	0	0	40.76955	0	62.36446
11.33694	12	30.8499	0	31.2488	0	18.23746	0	0	0	112	0	0	0	0	0	157.3152	0	173.5421
4.524879	12	33.01955	0	43.03453	0	33.84371	0	0	0	112	0	0	0	0	0	159.0574	0	205.4472

Figure 2: Screenshot of GLCM features of CT scans

SRE	LRE	GLN	RP	RLN	LGRE	HGRE
0.092084	27745.63	5550.907	0.040131	160.919	98.51569	5550.907
0.082718	29991.95	5090.79	0.036392	121.0962	95.43553	5090.79
0.0341	130901.1	2577.813	0.012596	403.02	32.97153	2577.813
0.138458	14599.06	7498.929	0.055611	374.6906	106.8623	7498.929
0.03171	108141.4	2729.341	0.014809	311.8537	47.24421	2729.341
0.138767	20335.25	6128.765	0.044769	291.4559	101.6224	6128.765
0.141707	18967.08	6487.141	0.047623	321.6141	103.2329	6487.141
0.156462	23312.26	5864.684	0.042656	355.7453	100.2908	5864.684
0.171217	23134.36	6024.701	0.043938	445.8189	101.1137	6024.701
0.133001	20985.48	6052.369	0.044159	278.443	101.251	6052.369
0.169197	17438.27	7193.411	0.053207	529.0182	105.8849	7193.411
0.179397	13963.29	8691.766	0.064941	777.7877	109.9712	8691.766
0.155459	22657.21	6078.142	0.044365	375.2838	101.3775	6078.142
0.033484	92961.16	2832.576	0.016068	190.0774	53.6104	2832.576
0.153764	22019.77	5994.185	0.043694	366.9176	100.9607	5994.185

Figure 3: Screenshot of GLRLM features of CT scans



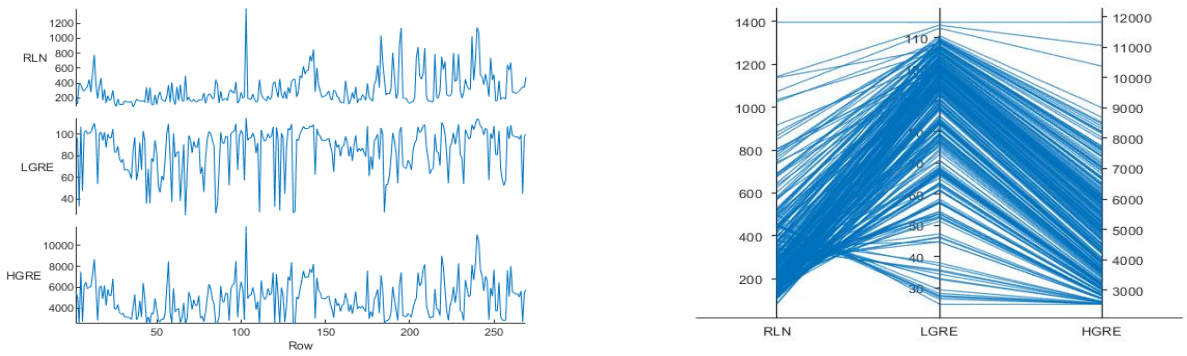


Figure 4: Stackedplot and parallelplot representation of GLCM and GLRLM features

The stackedplot graph visualizes the composition and comparison of the features variables in different y-axes and common x-axis. Parallelplot graph creates a parallel coordinates plot from the feature extraction file. Each line in the plot represents the features extracted for each CT scan image, and each coordinate variable in the plot corresponds to the feature variables.

Conclusion:

In this experimental approach we conclude that the Gray-Level Co-Occurance Matrix (GLCM) and Gray-level run-length matrix (GLRLM) methods are highly efficient for texture feature extraction of CT images of liver. The extracted values when fed in to the support vector machine classifier satisfactorily categorize the elements into different tumor classes as expected. The model best describes a healthy tissue and that of an unhealthy tissue in the liver organ. This model could be of great help in the decision making process of liver tumor diagnosis and treatment in the medical field.

References:

[1] Baron, R.,"Understanding and optimizing use of contrast material for CT of the liver",American Journal of Roentgenology , vol.163, 1994, pp323–331.
[2] Wan Nur Hafsha Wan Kairuddin and Wan Mahani Hafizah Wan Mahmud 2017,*IOP Conf. Ser.: Mater. Sci. Eng*. 226**,** 012136
[3] Rituparna Sarma and Yogesh Kumar Gupta (2021). A comparative study of new and existing segmentation techniques. *IOP Conf. Ser.: Mater. Sci. Eng*. 1022(1) ,012027
[4] Sertel, O., Lozanski, G., Shana'ah, A., &Gurcan, M. N. (2010). Computer-aided detection of centroblasts for follicular lymphoma grading using adaptive likelihood-based cell segmentation. IEEE Transactions on Biomedical Engineering, 57(10), 2613-2616.
[5] Mikhaylov, V. V., &Bakhshiev, A. V. (2017). The System for Histopathology Images Analysis of Spinal Cord Slices. Procedia Computer Science, 103, 239-243.
[6] Albregtsen, F., Nielsen, B., &Danielsen, H. E. (2000). Adaptive gray level run length features from class distance matrices. In Pattern Recognition, 2000. Proceedings. 15th International Conference on (Vol. 3, pp. 738-741). IEEE.
[7] Mohanty, A. K., Beberta, S., &Lenka, S. K. (2011). Classifying benign and malignant mass using GLCM and GLRLM based texture features from mammogram. International Journal of Engineering Research and Applications, 1(3), 687-693.
[8] Sastry, S. S., Kumari, T. V., Rao, C. N., Mallika, K., Lakshminarayana, S., &Tiong, H. S. (2012). Transition temperatures of thermotropic liquid crystals from the local binary gray level cooccurrence matrix. Advances in Condensed Matter Physics, 2012.
[9] Mohanaiah, P., Sathyanarayana, P., &GuruKumar, L. (2013). Image texture feature extraction using GLCM approach. International Journal of Scientific and Research Publications, 3(5), 1
[10] Nanni, L., Brahnam, S., Ghidoni, S., Menegatti, E., & Barrier, T. (2013). Different approaches for extracting information from the co-occurrence matrix. PloS one, 8(12), e83554.
[11] Vapnik, V., &Izmailov, R. (2017). Knowledge transfer in SVM and neural networks. Annals of Mathematics and Artificial Intelligence, 1-17.
[12] Rituparna Sarma and Yogesh Kumar Gupta (2022). Volume based segmentation and classification of liver tumor using marker controlled watershed transform. European Chemical Bulletin, 11(6), 806-812