# AUTOMATIC CLASSIFICATION OF ASTROCYTE CELLS FOR TRAUMA USING SPATIAL AND TRANSFORM DOMAIN FEATURES

**Mamta Kumari[1] Vishal Pareek[2]**
[1]M.Tech. Scholar, [2]Assoc. Professor,
[1,2]Department of Computer Science and Engineering,
Sri Ganganagar Engineering College, Sri Ganganagar

***Abstract— Machine learning has enabled the way for designing modern computing systems that works in similar way as a human expert. It is rapidly becoming an essential part of data processing applications in which the size of data is enormous and manual processing is not feasible. Machine learning techniques find critical applications in areas where traditional computing methods fail to compute. For example, a traditional computing algorithm can be designed to find weather a given number is greater than, or smaller than, or equal to a specific number. However, traditional computing fails to figure out how similar a given image is to a specific image or how similar (or dissimilar) is the content written in two separate paragraphs. In such cognitive computing applications, machine learning techniques play a very significant role. A machine learning algorithm, and a classifier based on it can be implemented using neural networks, Bayesian networks, Fuzzy logic or Support Vector Machine (SVM). In this paper, a SVM based classifier is implemented to study the correlation properties for Astrocyte cells and the variations caused in human brain in case of Trauma. The proposed machine takes as input, the microscopic images of the astrocyte cells and classifies it into either normal or in trauma state. Image preprocessing is performed to extract the feature set which serves as inputs for classification. This paper extends the work of Najiya Nasrin et.al. by including, apart from the spatial feature set, the frequency domain characteristics that carries significant information of image texture. The simulation is performed on R statistical package and it turns out that proposed methods gives a more accurate classification as compared to the benchmark techniques.***

***Keywords : Brain Cells, Feature Extraction, Classification, Machine Learning, SVM etc.***

## I Introduction

### 1.1 Introduction

Automatic Classification is one of the most important applications of Machine Learning. These techniques found tremendous applications in modern cognitive computing applications. It refers to the classification of a given number of items into a number of groups as per the classification requirements. Given a set of items, the classifier examine each item and labels it with a class label. In the simplest case, the classifier classifies each item of a given set into exactly one of the two specified classes, in which case it is known as a binary classifier. The task of classification is based on the examination and/or processing of certain parameters of the items known as the feature sets. Automatic classification is increasingly required in applications where the size of the data items to be classified is so large that the manual classification is time consuming and costly.

There exist two variants of classifiers corresponding to two popular branches of machine learning; viz supervised and unsupervised learning. Supervised Learning corresponds to the scheme in which a pre-classified data sample, called training data, is provided to the machine to train the machine. The machine analyses the data and the class labels to understand the relation between the two. The machine is then subjected to test data. The classification results given by the machine are also checked by comparing them to the manually classified data to compute the machine accuracy and precision. In case of binary classifiers, this is done by computing false positives and false negatives.

Another variant of classifier, which corresponds to unsupervised learning, is less popular and comes under clustering techniques. In this variant, there is no training data and the machine is directly assigned a data sample, with the objective to form logical clusters of the data items. One popular technique in this category is k-means clustering algorithm.

### 1.2 Image Classification

Image classification refers to the task of separating a given set of images into logical groups as required in the application. Understanding an image is a very simple task for human brain. In contrast, it takes considerably complex algorithms and techniques to design a program to interpret images. This is because for a computer system, an image is nothing more than a grid of pixels. Interpreting images requires the extraction of features from the images such as texture and edges. These features constitute meaningful information of the image. The more the number of informative features, the better is the design of a classifier for logical classification.

The intent of the classification process is to categorize all pixels in a digital image into one of several classes. This categorized data may then be used to produce thematic maps of the type of object present in an image. Normally, multispectral data are used to perform the classification and the spectral pattern present within the data for each image is used as the numerical basis for categorization. The objective of image classification is to identify and portray the features occurring in an image in terms of the object which these features actually represent.

Image classification is the most important application of digital image analysis. In almost all the image classification applications, the image is first subjected to pre-processing. It refers to atmospheric correction, noise removal, image transformation, main component analysis etc. Spatial domain feature extraction refers to extraction of features from the pixel values of the given image. Transform domain feature extraction refers to extraction of features after operating the image with DFT, DWT or similar transforms.

II .METHODOLOGY

A classifier is a program which classifies a given data set into logical classes. For example, a classifier can be designed to analyze the height and weight of students of a class to classify weather the student is normal, underweight or over-weight based on the pre-specified range for normal, over-weight and underweight as per the height of the candidate. The said classifier can be designed with traditional programming language using logical operators.

However, the real-life scenarios for classification of higher order data-sets cannot be done using traditional techniques. For example. given a specification of a flower in terms of its four components; say sepals, petals, stamen and pistils, it is difficult to design a program, using traditional programming, to tell whether the flower is mature flower, or a blossom. In such a case, machine learning techniques are required in which the machine is required to be trained on a pre-classified data set. This is illustrated in figure 1.1

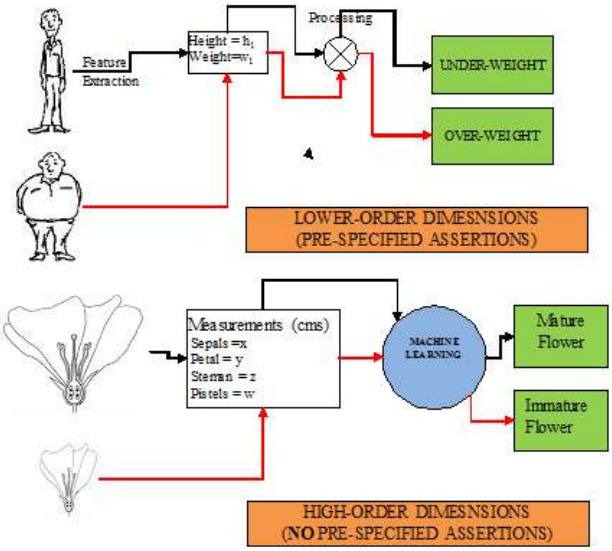



Fig 1.1 Illustration of Traditional and Machine Learning Classifier

In this paper, a model is proposed to classify the microscopic image of astrocyte cell for being in the normal state or in the trauma state. The image data set is obtained from the UCI machine learning repository. The extended set of features, including both the spatial and transform domain features are extracted for all the images. The pre-classified data set of images in normal and trauma state is given to the machine as training data. The given image samples are preprocessed to remove noise and then subjected to the subsequent feature extraction. Both the spatial and the transform domain features are extracted and a training data set is constructed. A support vector machine based classifier is designed and machine is trained on the manually classified data. Thereafter, test data is given to the machine. The classification results produced by the machine are checked for the computation of precision and accuracy of the machine.

This paper is organized as follows. Section 1 presents an overview of the subject matter and gives the problem statement and the approach for the research. Section 2 provides the detailed overview of various research papers that contributed to the subject. It also makes a background for the concepts and techniques presented in subsequent sections. Section 3 presents the proposed model for analysis of a review for image classification. Section 4 gives the simulation results and the plots for image classification using spatial and frequency domain features. Section 5 concludes the paper.

III .PROPOSED WORK

**3.1 Image Feature Extraction**

Stated mathematically, an image is a grid of pixel and an underlying hardware and software platform has no logical information about the image. The only stored information about the image is that it is a collection of numeric values in the database. Classification of a given set of images into logical groups based on the image logical content is called image classification. The task of classification of images is a complex computational problem for existing computer systems. This is due to the inherent difference in the computational characteristics of human brain and the computers. In this paper, the classification problem of images is investigated for classification of microscopic images of astrocyte cells for being in normal state or in trauma state. The feature extraction of images refers to the process of extracting important information (in specific context) from the image pertaining to the color, texture and edges so that these can be used as input for the automated classifier. Image feature extraction can be viewed as a form of dimensionality reduction in which a complex problem is broken down into simple problem for feasible investigation. The proposed images classifier takes as input, features from the GCLM of the original image as well as the edge extracted image. For extracting the edges, the Canny Edge detection technique is used as it is well accepted optimal technique for edge detection. However, comparative analysis has been made with other edge detection techniques. The implementation of the proposed model is done using MATLAB. The model for implementation first extract the feature set of the images which are manually classified into trauma and non-trauma. It then implements a two class classifier implemented using SVM.

**3.2 Proposed Work-Flow for Image Classification**

The proposed model can be depicted as shown:

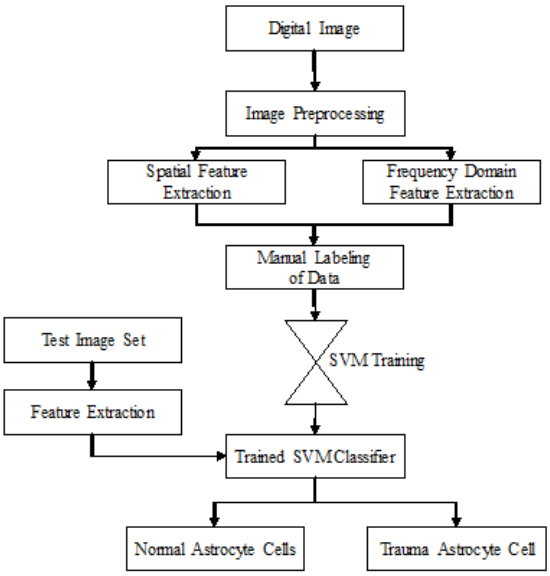



Fig 3.1 Workflow of the classification process

The above figure illustrates the generic aspects of the automated image classification technique. The SVM is trained over the manually labeled feature set comprising of GLCM features of the original image and the Canny edge detected image. The performance metrics are compared with those of the benchmark techniques implemented using classical edge detection techniques and conclusions are drawn. It is shown in subsequent sections that the proposed technique out-performs that of the Image Classification based on only GLCM of the original image with feature information.

**3.3 Gray Level Co-Occurrence Matrices**

The GLCM specifies the frequencies of different combinations of pixel brightness values (gray scale levels) in an image. GLCM describes the second order texture of the images. The first order texture consists of statistics from the pixel values of the image, like histogram, variance etc. The second order features of the image consist of the parameters which involves a pair of pixels from the image. Third and higher order texture measures are not commonly used due to the time and computational complexity. GLCM tabulates how the pairs of different levels of color brightness occur in an image. The computation of the GLCM is illustrated using the following example:

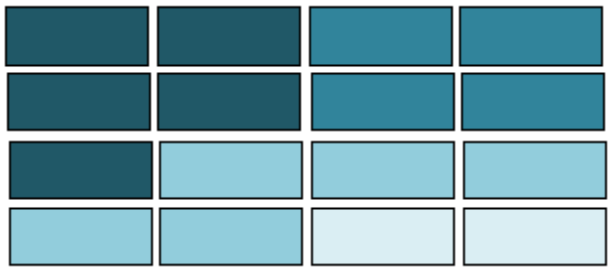


Fig 3.2 Illustrative Hypothetical Image

The pixel value matrix corresponding to the image shown in figure 3.2 is

| | | | |
|---|---|---|---|
| 0 | 0 | 1 | 1 |
| 0 | 0 | 1 | 1 |
| 0 | 2 | 2 | 2 |
| 2 | 2 | 3 | 3 |

Fig 3.3 Pixel Matrix corresponding to Figure 3.1

GLCM computes the relationship between two pixels at a time, called the reference and the neighborhood pixel. Suppose, one wants to tabulate the GLCM of adjacent pixels in which the neighbor pixel is the pixel in the right direction of the given pixel. For such GLCM, the tabulation is done in the following form:

| Neighbor Pixel Value→<br>Reference Pixel Value<br>↓ | 0 | 1 | 2 | 3 |
|---|---|---|---|---|
| 0 | 0,0 | 0,1 | 0,2 | 0,3 |
| 1 | 1,0 | 1,1 | 1,2 | 1,3 |
| 2 | 2,0 | 2,1 | 2,2 | 2,3 |
| 3 | 3,0 | 3,1 | 3,2 | 3,3 |

Fig 3.4 GLCM Construction

The top left corner consists of the entry 0,0. This will be replaced by the number of times the entry 0,0 appears in the image. The GLCM corresponding to the matrix in Fig 3.3 is shown in figure 3.5

| | | | |
|---|---|---|---|
| 2 | 2 | 1 | 0 |
| 0 | 2 | 0 | 0 |
| 0 | 0 | 3 | 1 |
| 0 | 0 | 0 | 1 |

Fig 3.5 Pixel Matrix corresponding to figure 3.2 (combination 1,0)

A different co-occurrence matrix exists for each spatial relationship, viz, above, next-to and diagonal. Typically, in the MATLAB implementation of GLCM, the matrix is converted to symmetrical so that the lower and upper diagonal values become the same. Moreover, the absolute values are replaced by the normalized values. Stated in another way, rather than the actual count of pixel pair combinations, the corresponding matrix element consists of the probability of occurrence of pixel pair.

The above illustration is shown for pixel pair in which the neighborhood pixel is the immediate right hand side pixel of the given reference pixel. However, GLCM construction, for second order, is possible for the pixel pair orientations as shown in figure 3.5.

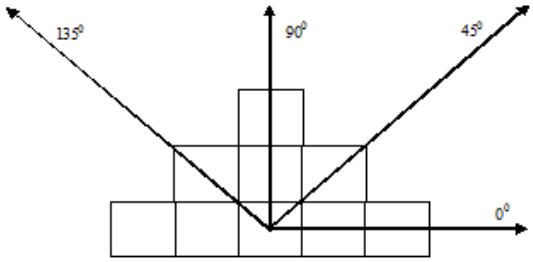



Fig 3.6 Pixel Pair Orientations for Second Order GLCM

**3.4 Feature Extraction**

As stated earlier, the GLCM defines a square matrix whose elements represents the Gray value $g_1$ distance from a fixed spatial location relationship in terms of size and direction to another gray value $g_2$.

Thus, GLCM can be expressed as:

$$P(i,j) = \frac{\#\{[(i_1,j_1),(i_2,j_2)] \in S | f(i_1,j_1) = g_1 \ \& \ f(i_2,j_2) = g_2\}}{\#S}$$

The most prominent features that can be extracted from the GLCM matrix include Angular Second Moment (ASM), Contrast (CON), Correlation (COR) and Entropy (ENT). These are defined as follows:

$$Angular\ Second\ Moment\ (ASM) = \sum_{i}^{N}\sum_{j}^{N} P(i,j)^2$$

$$Contrast\ (CON) = \sum_{i}^{N}\sum_{j}^{N} (i-j)^2 P(i,j)$$

$$Correlation\ (COR) = \frac{\sum_i^N \sum_j^N (i-\bar{x})(i-\bar{y})P(i,j)}{\sigma_x\, \sigma_y}$$

Where

$$\bar{x} = \sum_{i}^{N} i \sum_{j}^{N} P(i,j)$$

$$\bar{y} = \sum_{j}^{N} j \sum_{i}^{N} P(i,j)$$

$${\sigma_x}^2 = \sum_{i}^{N} (i-\bar{x})^2 \sum_{j}^{N} P(i,j)$$

$${\sigma_y}^2 = \sum_{j}^{N} (j-\bar{y})^2 \sum_{i}^{N} P(i,j)$$

And

$$Entropy\ (ENT) = -\sum_{i}^{N}\sum_{j}^{N} P(i,j) log\ P(i,j)$$

**3.5 Proposed SVM Classifier**

The proposed SVM Classifier is trained over a supervised learning approach. The feature set is extracted from the training sample of images and classifier is trained over the set. The test data set is provided to the trained classifier to obtain the classification results. The classification results are compared to those of the manually classified results. Precision, Accuracy and Recall are used as quality parameters to compare the results with those of the benchmark techniques. The proposed classification architecture is described as follows:

**Proposed Algorithm**

1. GLCM is obtained from the input set of astrocyte normal and astrocyte trauma microscopic images. Angular Second Moment, Contrast, Correlation and Entropy features are extracted from the images.

2. For the same input set, the edge detection techniques are applied and GLCM is applied to the set of edge detected images. All the above specified features, viz; Angular Second Moment, Contrast, Correlation and Entropy features are extracted from the edge detected images
3. The normal astrocyte cells are used as positive examples and the trauma astrocyte cells are used as negative examples to train the SVM.
4. When given an input image which is either normal sample or trauma sample, the machine returns with the result whether it is normal or not (means trauma state).
5. The two class sample is required as SVM is primarily a binary classifier. The proposed classifier actually computes the result whether or not the given sample is normal state. It has nothing to do with trauma state as these are supplied as negative examples. However, the converse training is possible, as the machine can be trained for trauma as positive examples. In this case, the machine returns with the result whether the given input sample is trauma state or not. In that case, the machine has nothing to do with whether or not the sample is normal or not.

The proposed technique is illustrated in figure 3.7. Thus, a total of 8 features are extracted from the image and a dataset of the feature values is tabulated as training data set. The accuracy, precision and recall of the machine is computed for Sobel, Perwitt and Canny Edge Detection techniques.

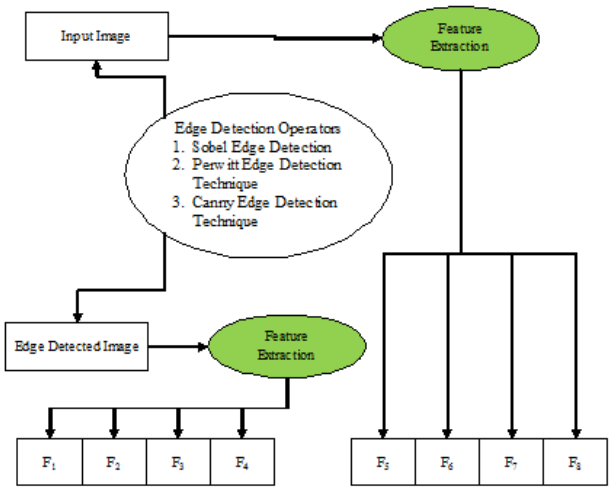



Fig 3.7 Proposed Architecture of Feature Set Generation

It is important to note that the extracted feature set form the dataset of only positive examples. The data set corresponding to the negative examples can be built using any technique to generate random values. In this paper, investigation has also been made on the accuracy of the results obtained from the training set of the edge detected images from all the three benchmark techniques for edge detection, viz; Sobel, Perwitt and Canny edge detection techniques.

The feature set of the normal and the edge detected images is used to train the SVM. The results are compared with those of the benchmark edge detected techniques. Precision and recall are used as quality metrics and the results are compared with those of the benchmark techniques.

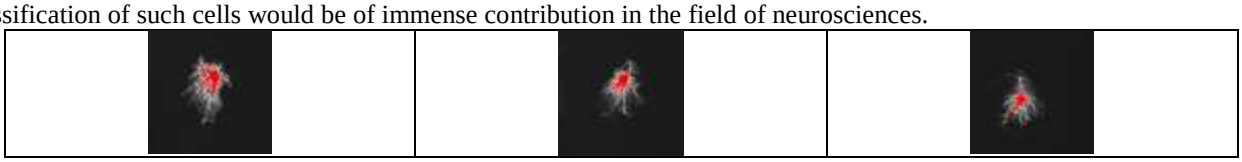
## IV ANALYSIS OF PROPOSED WORK

### 4.1 Data-Set of Astrocyte Cells

Astrocyte Cells are known to play a critical role in human brain. These cells perform a variety of tasks, from controlling the sensory organs to synaptic support. These cells are also responsible for emotions, memory and other aspects of human behavior. Astrocyte cells are also responsible for the creation of boundaries to compartmentalize the regions of the Human Brain. These boundaries are also created in the case of Trauma, when the brain is subjected to external impulse which may lead to temporary or permanent damage to certain cognitive, physical and/or psychological functions. In case of such impulses, Astrocyte cells typically create Laminar (smooth) boundaries which are induced by Trauma. The image dataset of Astrocyte cells is obtained from cell image data repository.

The dataset consists of images of common astrocyte cells along-with those astrocyte cells which represents such trauma induced boundaries. The principle investigation is the analysis of astrocyte cells to check whether or not these are parts of such boundaries. A sample of astrocyte cells is shown in figure 4.1. The sample consists of microscopic images of star shaped astrocyte cells inside the Dentate Gyrus of hippocampus. The hippocampus is thought to be the part of the brain ventricles responsible for the formation of memories, emotions and automatic nervous system. The sample is built through images of astrocytes in the adult rat Dentate Gyrus. These images were revealed with fluorescent tracer dyes and subsequently analyzed with respect to laminar boundaries demarcated by means of immune-labeling for the lamina-specific molecules EphA4 and Neural Cell Adhesion Molecule (N-CAM). The image sample is shown in figure 4.1. The automatic classification of such cells for the formation of Laminar Boundaries would be of immense help in treatment of patients in case of Trauma.

The grid shown in figure 4.1 tabulates the samples of both, ordinary Astrocyte Cells and those which belong to the laminar boundary formation. The red region corresponds to NCAM. The sample clearly indicates that it is difficult to differentiate two classes. However, an accurate classification of such cells would be of immense contribution in the field of neurosciences.



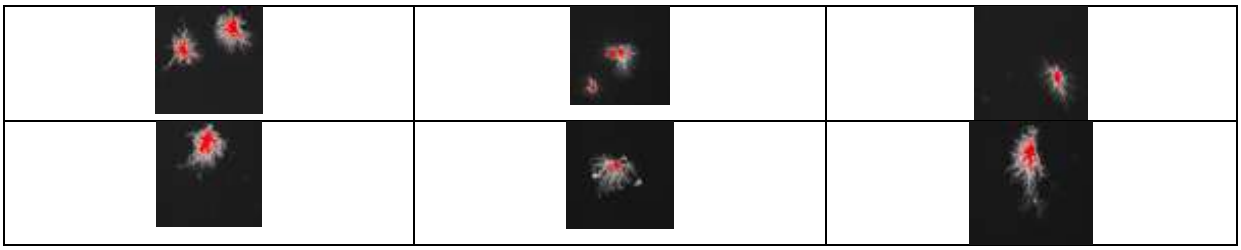


Fig 4.1 Image Samples of Astrocyte Cells

**4.2 Image Preprocessing Operations**

The image preprocessing comprises of Conversion to gray-scale and computation of the GLCM matrix. Taking any image specimen, the operation can be illustrated as shown:

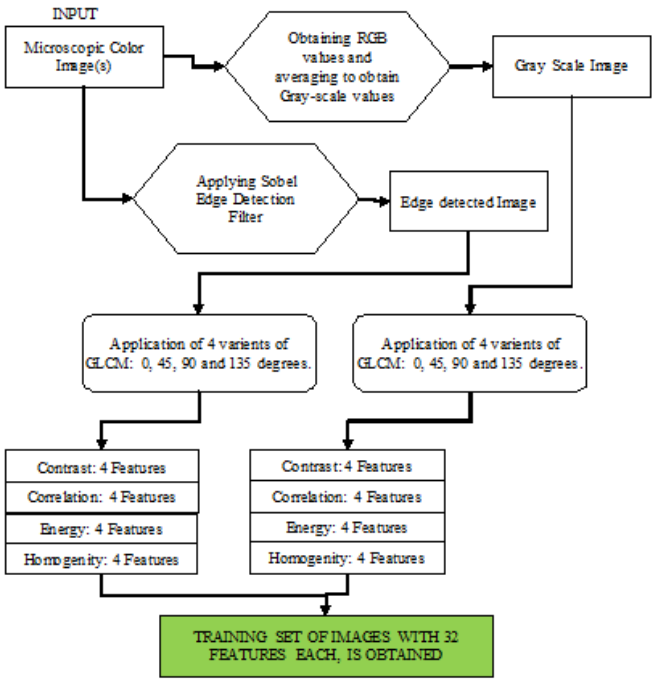



Fig 4.2 Schematic of Obtaining the Feature Set of Images (For Both Positive and Negative Examples)

A training set of 2000 images has been constructed out of which half of the images belong to normal astrocyte and other half corresponds to those that belong to Trauma state. The training set, consisting of a total of 32 features of every image is tabulated. This set is manually classified and labeled as the image data repository itself has labeled images, those that belongs to normal astrocyte and those that belongs to Trauma astrocytes. A sample of the record of the normal astrocyte cells is shown in Table 4.1

TABLE 4.1
FEATURE SET SAMPLE OF NORMAL ASTROCYRE (Sample 10 Records)

| | | | | | | | | |
|---|---|---|---|---|---|---|---|---|
| 1. | 0.027985873 | 0.042252442 | 0.028433066 | 0.042574132 | 0.96374787 | 0.945339845 | 0.963167674 | |
| | 0.944923689 | 0.57815257 | 0.567650178 | 0.578326183 | 0.567696822 | 0.986007063 | 0.978873779 | |
| | 0.985783467 | 0.978712934 | 0.027985873 | 0.042252442 | 0.028433066 | 0.042574132 | 0.96374787 | |
| | 0.945339845 | 0.963167674 | 0.944923689 | 0.57815257 | 0.567650178 | 0.578326183 | 0.567696822 | |
| | 0.986007063 | 0.978873779 | 0.985783467 | 0.978712934 | | | | |
| 2. | 0.019561827 | 0.02482757 | 0.019657381 | 0.026076034 | 0.971646298 | 0.964074423 | 0.971507709 | |
| | 0.962267892 | 0.75645336 | 0.753491133 | 0.756271437 | 0.753161962 | 0.990305722 | 0.987790462 | |
| | 0.990268137 | 0.987186655 | 0.019561827 | 0.02482757 | 0.019657381 | 0.026076034 | 0.971646298 | |
| | 0.964074423 | 0.971507709 | 0.962267892 | 0.75645336 | 0.753491133 | 0.756271437 | 0.753161962 | |
| | 0.990305722 | 0.987790462 | 0.990268137 | 0.987186655 | | | | |
| 3. | 0.022462849 | 0.029507393 | 0.02225263 | 0.029423141 | 0.958203237 | 0.945188895 | 0.95863634 | |
| | 0.945345397 | 0.75771384 | 0.753059054 | 0.756944088 | 0.753019567 | 0.988768576 | 0.985287153 | |
| | 0.988878781 | 0.985339491 | 0.022462849 | 0.029507393 | 0.02225263 | 0.029423141 | 0.958203237 | |
| | 0.945188895 | 0.95863634 | 0.945345397 | 0.75771384 | 0.753059054 | 0.756944088 | 0.753019567 | |
| | 0.988768576 | 0.985287153 | 0.988878781 | 0.985339491 | | | | |
| 4. | 0.015774064 | 0.021633649 | 0.0170392 | 0.021763857 | 0.968502173 | 0.956877102 | 0.96597611 | |
| | 0.956617555 | 0.796904767 | 0.79345666 | 0.79589423 | 0.793058845 | 0.992112968 | 0.989193388 | |
| | 0.9914804 | 0.989148709 | 0.015774064 | 0.021633649 | 0.0170392 | 0.021763857 | 0.968502173 | |
| | 0.956877102 | 0.96597611 | 0.956617555 | 0.796904767 | 0.79345666 | 0.79589423 | 0.793058845 | |
| | 0.992112968 | 0.989193388 | 0.9914804 | 0.989148709 | | | | |
| 5. | 0.041157045 | 0.06111726 | 0.039112188 | 0.06069217 | 0.950726153 | 0.926889493 | 0.953215218 | |
| | 0.927397778 | 0.388603091 | 0.373360535 | 0.389973486 | 0.373662148 | 0.979441862 | 0.969456689 | |

0.980454098 0.969669234 0.041157045 0.06111726 0.039112188 0.06069217 0.950726153
0.926889493 0.953215218 0.927397778 0.388603091 0.373360535 0.389973486 0.373662148
0.979441862 0.969456689 0.980454098 0.969669234

6. 0.036318187 0.04998832 0.039517337 0.050915093 0.944411043 0.923607331 0.939514219
0.922191026 0.659346272 0.649727304 0.657049403 0.649291302 0.981922446 0.9752203
0.980358544 0.974751807 0.036318187 0.04998832 0.039517337 0.050915093 0.944411043
0.923607331 0.939514219 0.922191026 0.659346272 0.649727304 0.657049403 0.649291302
0.981922446 0.9752203 0.980358544 0.974751807

7. 0.02456886 0.03586077 0.026900379 0.034229342 0.943109136 0.917059174 0.937678034
0.920832432 0.75039805 0.74215946 0.748921265 0.742806919 0.98771557 0.982084934
0.98654981 0.982895541 0.02456886 0.03586077 0.026900379 0.034229342 0.943109136
0.917059174 0.937678034 0.920832432 0.75039805 0.74215946 0.748921265 0.742806919
0.98771557 0.982084934 0.98654981 0.982895541

8. 0.042433647 0.061289594 0.044145976 0.061021519 0.949010095 0.926417069 0.946922645
0.926738913 0.595139782 0.582054373 0.594865465 0.582280653 0.97881885 0.969513495
0.977988167 0.969601577 0.042433647 0.061289594 0.044145976 0.061021519 0.949010095
0.926417069 0.946922645 0.926738913 0.595139782 0.582054373 0.594865465 0.582280653
0.97881885 0.969513495 0.977988167 0.969601577

9. 0.009096747 0.011944654 0.00857311 0.011944654 0.928621996 0.906413909 0.93270808
0.906413909 0.8635522 0.860576883 0.864109483 0.860576107 0.995451627 0.994027673
0.995713445 0.994027673 0.009096747 0.011944654 0.00857311 0.011944654 0.928621996
0.906413909 0.93270808 0.906413909 0.8635522 0.860576883 0.864109483 0.860576107
0.995451627 0.994027673 0.995713445 0.994027673

10. 0.027137353 0.035439509 0.027274951 0.035998637 0.954595668 0.940806911 0.954365449
0.939873024 0.774210162 0.770459177 0.773914365 0.770223838 0.986533248 0.982938944
0.986520507 0.982598106 0.027137353 0.035439509 0.027274951 0.035998637 0.954595668
0.940806911 0.954365449 0.939873024 0.774210162 0.770459177 0.773914365 0.770223838
0.986533248 0.982938944 0.986520507 0.982598106

Table 4.2 shows the sample 10 records of feature set of images corresponding to Trauma astrocyte.

TABLE 4.2
FEATURE SET SAMPLE OF TRAUMA ASTROCYRE (Sample 10 Records)

1. 0.118085708 0.15701916 0.124296722 0.155399221 0.759816746 0.680855148 0.747418229
0.684149979 0.503142335 0.476031741 0.49837053 0.477058431 0.940992819 0.9216385
0.93787712 0.922422938 0.118085708 0.15701916 0.124296722 0.155399221 0.759816746
0.680855148 0.747418229 0.684149979 0.503142335 0.476031741 0.49837053 0.477058431
0.940992819 0.9216385 0.93787712 0.922422938

2. 0.081591701 0.110236251 0.087718628 0.111863849 0.865461875 0.818161305 0.85538177
0.815476168 0.415328673 0.394888179 0.410566515 0.393731331 0.959362132 0.945458874
0.95632415 0.944706349 0.081591701 0.110236251 0.087718628 0.111863849 0.865461875
0.818161305 0.85538177 0.815476168 0.415328673 0.394888179 0.410566515 0.393731331
0.959362132 0.945458874 0.95632415 0.944706349

3. 0.045686307 0.0664711 0.045865949 0.066601307 0.804396644 0.715558259 0.803806473
0.715001076 0.736212373 0.717908015 0.735859798 0.717808893 0.977167039 0.96676445
0.977072122 0.966704452 0.045686307 0.0664711 0.045865949 0.066601307 0.804396644
0.715558259 0.803806473 0.715001076 0.736212373 0.717908015 0.735859798 0.717808893
0.977167039 0.96676445 0.977072122 0.966704452

4. 0.00503761 0.006793785 0.005152275 0.006908675 0.941427714 0.92115568 0.940094506
0.919822348 0.908985064 0.907089275 0.908871565 0.906975958 0.997481195 0.996603107
0.997423863 0.996545663 0.00503761 0.006793785 0.005152275 0.006908675 0.941427714
0.92115568 0.940094506 0.919822348 0.908985064 0.907089275 0.908871565 0.906975958
0.997481195 0.996603107 0.997423863 0.996545663

5. 0.011722572 0.017436361 0.013408146 0.017053397 0.968879676 0.953789801 0.964404159
0.954804741 0.783897744 0.780374757 0.783181147 0.780554529 0.994138714 0.99128182
0.993295927 0.991473302 0.011722572 0.017436361 0.013408146 0.017053397 0.968879676
0.953789801 0.964404159 0.954804741 0.783897744 0.780374757 0.783181147 0.780554529
0.994138714 0.99128182 0.993295927 0.991473302

6. 0.180841793 0.2358983 0.182611454 0.233137128 0.825834379 0.772684733 0.824092499
0.775346035 0.255873598 0.23084585 0.254918967 0.231941581 0.910205938 0.883843123
0.909234472 0.88516754 0.180841793 0.2358983 0.182611454 0.233137128 0.825834379
0.772684733 0.824092499 0.775346035 0.255873598 0.23084585 0.254918967 0.231941581
0.910205938 0.883843123 0.909234472 0.88516754

7. 0.012379984 0.017260197 0.012521404 0.01669341 0.874668658 0.825585447 0.873242378
0.831312838 0.889023398 0.884105532 0.888882784 0.884653159 0.993810008 0.991369901
0.993739298 0.991653295 0.012379984 0.017260197 0.012521404 0.01669341 0.874668658
0.825585447 0.873242378 0.831312838 0.889023398 0.884105532 0.888882784 0.884653159
0.993810008 0.991369901 0.993739298 0.991653295

8. 0.015135763 0.019086937 0.01511283 0.019515857 0.933454302 0.916192801 0.933519145
0.914309493 0.886038481 0.883312108 0.885941648 0.883101149 0.992467792 0.990568868
0.992489451 0.990359514 0.015135763 0.019086937 0.01511283 0.019515857 0.933454302
0.916192801 0.933519145 0.914309493 0.886038481 0.883312108 0.885941648 0.883101149
0.992467792 0.990568868 0.992489451 0.990359514
9. 0.027511925 0.032371965 0.029426829 0.03274727 0.928348082 0.915841353 0.923360344
0.914865658 0.829811102 0.827379265 0.828698975 0.827572649 0.986529426 0.984528884
0.985684091 0.984382081 0.027511925 0.032371965 0.029426829 0.03274727 0.928348082
0.915841353 0.923360344 0.914865658 0.829811102 0.827379265 0.828698975 0.827572649
0.986529426 0.984528884 0.985684091 0.984382081
10. 0.081196434 0.107570064 0.084645078 0.104967314 0.949679713 0.9333363 0.947552081
0.934949338 0.323038701 0.310617987 0.321400989 0.31181397 0.959431931 0.946537256
0.957761876 0.947813283 0.081196434 0.107570064 0.084645078 0.104967314 0.949679713
0.9333363 0.947552081 0.934949338 0.323038701 0.310617987 0.321400989 0.31181397
0.959431931 0.946537256 0.957761876 0.947813283

It is important to note that the feature set of both the normal and the trauma astrocyte cells consists of 32 features each which are explained as shown in table 4.3

TABLE 4.3
FEATURE SET COMPUTATION: GREYSCALE (COLORED AVERGE) IMAGE

| Orientation of Reference Pixel/Parameter | Contrast | Correlation | Energy | Homogeneity |
|---|---|---|---|---|
| **Orientation 0 degrees** | Feature #1 | Feature #5 | Feature #9 | Feature #13 |
| **Orientation 45 degrees** | Feature #2 | Feature #6 | Feature #10 | Feature #14 |
| **Orientation 90 degrees** | Feature #3 | Feature #7 | Feature #11 | Feature #15 |
| **Orientation 135 degrees** | Feature #4 | Feature #8 | Feature #12 | Feature #16 |

Another set of 16 Features is obtained from the Edge Detected Image which is obtained after Edge detection over Gray-scale using Sobel Filter. Thus, a total of 32 features are obtained from each image.

With a repository of 2000 images, in which half of the image corresponds to normal astrocyte and other half corresponds to that of Trauma astrocytes, a dataset of 2000 records each having 32 features set is constructed.

The proposed classifier and the workflow proceed as follows:

1. Obtain image dataset of 1000 images each of normal astrocyte and Trauma astrocyte from open microscopic image repository.
2. Obtain the 32 Feature set of all the 2000 images obtained and label them as Class A and Class B corresponding to Normal and Trauma sets.
3. Divide each of the set of 1000 records into 7:3 proportions in which 700 records belongs to Training Set and the rest 300 belongs to the Test Set.
4. Implement an SVM classifier and train it using training data set.
5. Test the machine over Test data set to compute machine accuracy and precision.

The dataset is constructed using MATLAB and the SVM is implemented in R statistical package. The quality metric of the proposed classifier are precision and recall which are explained as follows:

Given a total of N images are provided as input to a Binary Classifier trained on a particular set, let the number of records classified by the classifier as of class A be $N_A$, and those of set be $N_B$ . Clearly,

$$N = N_A + N_B$$

Also, let the total number of images which actually belongs to class A, among those of $N_A$ is $N_a$.

The precision and the recall can now be described as follows:

$$Precision = \frac{Number\ of\ Relevant\ images\ Retrieved}{Number\ of\ Retrieved\ images} = \frac{N_a}{N_A}$$

Thus

$$Precision = \frac{tp}{tp + fp}$$

Here, tp refers to true positive, fp refers to false positive and fn refers to false negative under standard definitions.

The Recall is the fraction of relevant images retrieved from among all the images that actually belongs to the category of particular class.

$$Recall = \frac{Number\ of\ Relevant\ Images\ Retrieved}{Number\ of\ Images\ which\ belongs\ to\ the\ class}$$

Thus

$$Recall = \frac{tp}{tp + fn}$$

**4.3 Simulation Outcomes**

The SVM implementation has been done on the GLCM matrix consisting of a Feature Set of 32 features. As indicated in previous section, 16 of the features belongs to averaged Gray-scale image and the other 16 belongs to the edge detected image through SOBEL filter. The SVM model details are as shown:

Call:
svm.default(x = x, y = y)
Parameters:
SVM-Type: eps-regression
SVM-Kernel: radial

cost: 1
gamma: 0.03125
epsilon: 0.1
Number of Support Vectors: 46

Out of a sample of 40 images, 14 are considered as training sample and 26 are considered as test samples. The prediction table from the SVM is as shown:

TABLE 4.4
PREDICTION RESULTS

| S. No. | Test Indexes | Predicted Class | Actual Class | Match =? |
|---|---|---|---|---|
| 1 | 20 | 1 | 0 | FALSE |
| 2 | 75 | 1 | 1 | TRUE |
| 3 | 40 | 0 | 0 | TRUE |
| 4 | 59 | 0 | 1 | FALSE |
| 5 | 46 | 1 | 1 | TRUE |
| 6 | 36 | 0 | 0 | TRUE |
| 7 | 41 | 1 | 1 | TRUE |
| 8 | 38 | 0 | 0 | TRUE |
| 9 | 39 | 1 | 0 | FALSE |
| 10 | 35 | 1 | 0 | FALSE |
| 11 | 3 | 0 | 0 | TRUE |
| 12 | 62 | 1 | 1 | TRUE |
| 13 | 28 | 1 | 0 | FALSE |
| 14 | 57 | 1 | 1 | TRUE |
| 15 | 72 | 1 | 1 | TRUE |
| 16 | 30 | 0 | 0 | TRUE |
| 17 | 1 | 0 | 0 | TRUE |
| 18 | 49 | 1 | 1 | TRUE |
| 19 | 21 | 1 | 0 | FALSE |
| 20 | 8 | 0 | 0 | TRUE |
| 21 | 11 | 0 | 0 | TRUE |
| 22 | 26 | 1 | 0 | FALSE |
| 23 | 13 | 0 | 0 | TRUE |
| 24 | 6 | 0 | 0 | TRUE |
| 25 | 14 | 0 | 0 | TRUE |
| 26 | 51 | 1 | 1 | TRUE |

Out of Total of 26 test samples, 7 samples are found incorrect matched and 19 samples are correctly predicted by the machine. The confusion matrix of the simulation model can be outlined as follows:

TABLE 4.5
CONFUSION MATRIX

| Actual → Predicted ↓ | Normal (0) | Astrocyte (1) |
|---|---|---|
| Normal (0) | 11 | 1 |
| Astrocyte (1) | 6 | 8 |

The confusion matrix using only the averaged gray-scale image, without using any detection filters gives the prediction results and confusion matrix indicated in table 4.6 and 4.7

TABLE 4.6
PREDICTION RESULTS USING ONLY AVERAGED GRAYSCALE IMAGE

| S. No. | Test Indexes | Predicted Class | Actual Class | Match =? |
|---|---|---|---|---|
| 1 | 20 | 1 | 0 | FALSE |
| 2 | 75 | 1 | 1 | TRUE |
| 3 | 40 | 0 | 0 | TRUE |
| 4 | 59 | 0 | 1 | FALSE |
| 5 | 46 | 1 | 1 | TRUE |

| | | | | |
|---|---|---|---|---|
| 6 | 36 | 0 | 0 | TRUE |
| 7 | 41 | 0 | 1 | FALSE |
| 8 | 38 | 0 | 0 | TRUE |
| 9 | 39 | 1 | 0 | FALSE |
| 10 | 35 | 1 | 0 | FALSE |
| 11 | 3 | 0 | 0 | TRUE |
| 12 | 62 | 0 | 1 | FALSE |
| 13 | 28 | 1 | 0 | FALSE |
| 14 | 57 | 1 | 1 | TRUE |
| 15 | 72 | 1 | 1 | TRUE |
| 16 | 30 | 1 | 0 | FALSE |
| 17 | 1 | 0 | 0 | TRUE |
| 18 | 49 | 1 | 1 | TRUE |
| 19 | 21 | 1 | 0 | FALSE |
| 20 | 8 | 0 | 0 | TRUE |
| 21 | 11 | 0 | 0 | TRUE |
| 22 | 26 | 1 | 0 | FALSE |
| 23 | 13 | 1 | 0 | FALSE |
| 24 | 6 | 0 | 0 | TRUE |
| 25 | 14 | 0 | 0 | TRUE |
| 26 | 51 | 1 | 1 | TRUE |

TABLE 4.7
CONFUSION MATRIX-BENCHMARK TECHNIQUE USING ONLY AVERAGED GRAYSCALE IMAGE

| Actual → Predicted ↓ | Normal (0) | Astrocyte (1) |
|---|---|---|
| Normal (0) | 09 | 3 |
| Astrocyte (1) | 7 | 6 |

The comparison of the resuts of the proposed and the benchmark techniques is reflected in the confusion matrix in table 4.2 and 4.4 respectively. The same is also indicated in table 4.8 and Figure 4.3.

TABLE 4.8
COMPARISON WITH BENCHMARK TECHNIQUES

| Cell Type | Total | Benchmark Approach | Proposed Approach |
|---|---|---|---|
| Normal | 17 | 9 | 11 |
| Astrocyte | 9 | 6 | 8 |

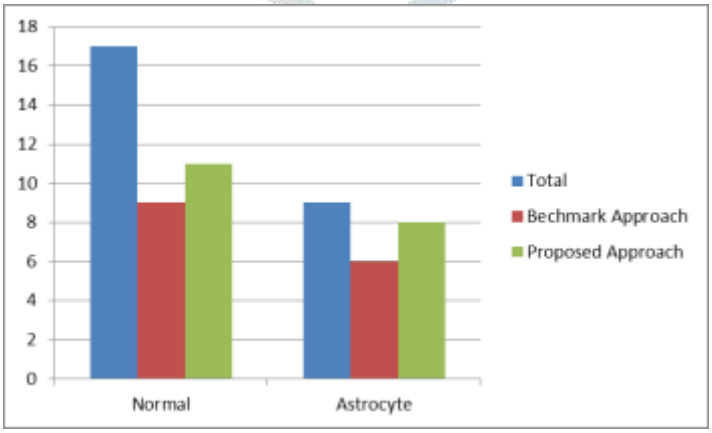



Figure 4.3 Comparisons with the Base Approach

The results obtained clearly indicate that the proposed approach outperforms that of the benchmark approach using only the gray-scale features. Section 4 gives a brief overview of the results and also indicates the proposed future scope of the work.

## V CONCLUSION AND FUTURE SCOPE

### 5.1 Conclusion

In this paper a technique is proposed for automatic classification of the microscopic images of the astrocyte cells in the human brain. The proposed technique is simulated using MATLAB and R statistical package. The results obtained outperform those of the benchmark

techniques. The research techniques presented here made use of the image processing operations to obtain a large class of the feature set of the image database. SOBEL filter is used for edge detection and another set of same number of feature vectors are obtained for edge detected image as those which are developed for averaged gray-scale image. A total of 32 feature vectors are used in SVM implementation which is much greater than any other proposed technique in the literature so far.

**5.2 Future Scope**

In future, research will be carried out on evaluation of the performance of the classification as a function of edge detection filters as well as the kernel function of the SVM. There exists a number of edge detection filters, the most popular of them includes the Canny Filter, Sobel Filter etc. In the similar way, there are a number of kernel functions available for SVM implementation.

An Extended Feature set can be obtained using two or more edge detection filters. The performance of the classification can be improved using both the extended feature set and the suitable kernel function.

REFERENCES


**[1]** Sagar Patil, Anjali Chandavale, Bharati Dixit, *Recognition of Plant Disease*, SPPU, Pune, iPGCON-2015.

**[2]** Vadhri Suryanarayana, Dr. M.V.L.N. Raja Rao, Dr. P. Bhaskara Reddy and Dr. G. Ravindra Babu,*Image Retrieval System Using Hybrid Feature Extraction Technique*, International Journal of Computer Science and Information Technology (IJCSIT), Vol 4, No 1, Feb 2012.

**[3]** Simona E. Grigorescu, Nicolai Petkov and Peter Kruizinga,*Comparison of Texture Features Based on Gabor Filters*, IEEE Transactions On Image Processing, Vol. 11, No. 10, October 2002 .

**[4]** Aswini Kumar Mohanty, Saroj Kumar Lenka and Swapnasikta Beberta, *Classifying Benign and Malignant Mass using GLCM and GLRLM based Texture Features from Mammogram* , International Journal of Engineering Research and Applications (IJERA), Vol. 1, Issue 3, pp.687- 693.

**[5]** A. D. Belsare and M. M. Mushrif, *Histopathological Image AnalysisUsing Image Processing Techniques: An Overview*, Signal & Image Processing : An International Journal (SIPIJ), Vol.3, No.4, August 2012.

**[6]** Neelima Bagri and Punit Kumar Johari *A Comparative Study on Feature Extraction using Texture and Shape for Content Based Image Retrieval*, International Journal of Advanced Science and Technology, Vol. 80 (2015), pp.41-52.

**[7]** Athira Krishnan, Sreekumar K .*A Survey on Image Segmentation and Feature Extraction Methods for Acute Myelogenous Leukemia Detection in Blood Microscopic Images*, (IJCSIT) International Journal of Computer Science and Information Technologies, Vol. 5 (6) , 2014, 7877-7879.

**[8]** Dr. H.B.Kekre, Sudeep D. Thepade, Tanuja K. Sarode and Vashali Suryawanshi.*Image Retrieval using Texture Features extracted from GLCM, LBG and KPE*, International Journal of Computer Theory and Engineering, Vol. 2, No. 5, October, 2010.

**[9]** Nikita Orlov , Lior Shamir , Tomasz Macura, Josiah Johnston , D. Mark Eckley and Ilya G. Goldberg. *WND-CHARM: Multi-purpose image classification using compound image transforms*, Pattern Recognition Letters 29 (2008) 16841693.

**[10]** Bino Sebastian V, A. Unnikrishnan and Kannan Balakrishnan. *Gray Level Co-Occurrence Matrices: Generalisation And Some New Features*, International Journal of Computer Science, Engineering and Information Technology (IJCSEIT), Vol. 2, No. 2, April 2012 .