Classification of Childhood Medulloblastoma into W.H.O. defined multiple subtypes based on Textural Analysis

Daisy Das[a], Lipi B. Mahanta[a]*, Shabnam Ahmed[b], Basanta K. Baishya[c]

[a]Institute of Advanced Study in Science and Technology, Guwahati, India;

[b]Guwahati Neurological Research Center, Sixmile, Guwahati, India;

[c]Gauhati Medical College and Hospital, Bhangagarh, Guwahati, India

*Department of CCNS, IASST, Vigyan path, Paschim Boragaon

*email@lbmahanta@iasst.gov.in

*contact@ +91 9706041892

ABSTRACT

Childhood medulloblastoma is a case of a childhood brain tumor that requires close attention due to the low survival rate. Effective prognosis depends a lot on accurate detection of its subtype. The present study proposes a texture-based computer-aided categorization of childhood medulloblastoma samples. According to the World Health Organization (W.H.O), it has four subtypes (desmoplastic, classic, nodular, and large). Classification is done in two levels: i) normal and abnormal ii) its four subtypes. The system is evaluated on indigenous patient samples collected from the region. The main objective of database generation is to create a dataset of childhood medulloblastoma samples since there exists no available benchmark dataset. The proposed framework for automated classification is based on the architectural property and the distribution of cells. Five texture features were extracted for the feature set viz: Gray Level Co-occurrence Matrix (GLCM), Gray Level Run Length Matrix (GRLM), First-order histogram features (HOGL), Local Binary Pattern (LBP) and Tamura features. The performance of

This article has been accepted for publication and undergone full peer review but has not been through the copyediting, typesetting, pagination and proofreading process, which may lead to differences between this version and the Version of Record. Please cite this article as doi: 10.1111/jmi.12893

each feature set was evaluated, both individually and in combinations, using five different classifiers. 5-fold cross-validation was used for training and testing the dataset. Experiments on both individual feature sets and combinations (best-2, best-3, best-4, all-5) of feature sets were evaluated based on the accuracy of performance. It was revealed that the combined best-4 feature set resulted in the highest accuracy of 91.3%. The Precision, Recall, and Specificity were 0.913, 0.913, and 0.97, respectively. Significantly, it implied that the all-5 feature set is not necessary to have a useful classification. Feature reduction by PCA resulted in increased accuracy of 96.7%.

ADDITION OF SECOND ABSTRACT -- LAY DESCRIPTION: Childhood medulloblastoma is a case of childhood brain tumor that requires high attention due to a low survival rate. Effective prognosis depends a lot on accurate detection of its subtype. The present study proposes a texture-based computer-aided categorization of childhood medulloblastoma samples. According to the World Health Organization (W.H.O), it has four subtypes (desmoplastic, classic, nodular, and large). Classification is done in two levels: i) normal and abnormal ii) its four subtypes. The system is evaluated on indigenous patient samples collected from the region. The main objective of database generation is to create a dataset of childhood medulloblastoma samples since there exists no available benchmark dataset. The proposed framework is a model for the automatic classification of the samples. The tissue samples obtained post-operation by doctors are converted into images, and then necessary algorithms are applied so that certain features describing each group of the image are known and studied for classification. Later these images are classified using the image features into the subtypes of abnormal samples.

Keywords: Medulloblastoma, Classification, Texture, Subtype, Machine learning, PCA

INTRODUCTION

Medulloblastoma is the most common malignant tumor in children accounting for most of all brain tumors. This densely cellular, midline cerebellar tumor arises over the roof of the fourth ventricle. The microscopic view of the tissue level displays a sheets-like arrangement of tightly packed cells, with large dark size nuclei. According to Polednak and Flannery[1], it constitutes approximately 20% of all childhood primary nervous system tumors. It has four subtypes[2] based on histology: classic, nodular, large cell, and desmoplastic. Each subtype has different architectural information and diagnosis. The large cell variant of Medulloblastoma resembles the classic pattern. It is architecturally the most prevalent and is highly anaplastic. As far as the molecular classification is concerned, medulloblastoma subtypes include WNT-activated, SHH-activated, and non-WNT/non-SHH (group 3 and group 4)[3]. The treatment procedure depends on subtype classification, as the degree of aggressiveness, differs from subtype to subtype. An improved 2- and 5-yr survival rate with the correct diagnosis has been reported[4] for childhood medulloblastoma. Providing an integrated diagnosis is of clinical value as both the histological and molecular variants have distinct therapeutic and prognostic implications[5].

CONTRIBUTION

The manual classification can vary, depending on individual pathologist experience and knowledge. This variation can be overcome by the use of computer-aided diagnostic tools that have been used for decades for better evaluations in medical and clinical analysis. Most studies relating to medulloblastoma[6-8] classified the tissue samples as either anaplastic or nonanaplastic. However, medulloblastoma tumor has four subtypes, and to the best of our knowledge, no attempt has been made to classify the tumor into these subtypes using computer-assisted

techniques to date. Therefore, we attempt to classify medulloblastoma into its subtypes, as declared by W.H.O. Also, the non-availability of public benchmark data is an added disadvantage responsible for lack of such work. The strength of this study is in the creation of a dataset of indigenous samples collected from the region. The generation of the dataset is based on the diagnosis provided by the clinicians and pathologists. The up-gradation of this database is an ongoing process as the study is in progress.

TEXTURAL ANALYSIS

From the histopathology point of view, analysis can be of 4 levels: morphology, color, texture, and sparse features. The pathologist performs the diagnosis both at 10x (architectural) and 40x (cellular) microscopic view for two-level observation of the tissue sample. The color-based feature helps us to understand the cell chromaticity and is done in 40x microscopic view. The cell morphology-based methods of digital histological analysis, also done at 40x view, extract the cells from the smear images and studies the various morphology of the cell associated with its shape, size, and structural features, for classification into different subgroups. However, it cannot depict the distribution of cells, i.e., the architectural level information of the tissue type, which is a vital diagnosis aspect of childhood medulloblastoma samples. The architectural view is a low power 10x microscopic view that a pathologist uses to check the distribution of cells. Texture-based classification[9] plays a significant role in the study of cell distribution and variation at the architectural level of tissue. Over the decades, a computer-based diagnosis has become very popular in the field of medical analysis. Therefore, it can be effectively used for a quick diagnosis of such tumor subtypes. A study by Awwad et al[10] has reported the potential for automated classification of tumor type and subtype in pediatric posterior fossa tumors based on standard MRI applying texture analysis. Sparse textures[11] (also known as tiled textures or mega-textures) are textures that are too huge to fit entirely in the visual memory. Still, since our

images were smaller and adequately handled by our machine Graphics Processing Unit (GPU), we did not opt for sparse based texture features.

METHOD AND MATERIAL

Informed consent and ethical clearance

The study conformed to the principles and guidelines of Rule 122DD, Drugs and Cosmetics Rule, 1945 of India and was approved by the Ethics Committee (Registration number ECR/248/Indt/AS/2015) of the Institute of Advanced Study in Science and Technology, Guwahati (approval number IEC(HS)/IASST/1082/2017-18/1), India. All participants provided written informed consent.

Implementation

The experiments were implemented in Matlab (R2016b, Mathworks, and Natick, MA, The USA) and All-in-one HP pc (2.70 GHz, Intel Core i5, 4 GB Ram). Also, built-in Matlab function, customized routines, and a portion of available source code [13-16] were used for the study.

Model

The block diagram (**Fig. 1**) describes the execution model of our work. The tissue blocks, collected from Guwahati Medical College and Hospital (GMCH) (step 1) and stained at Ayursundra

Healthcare Pvt. Ltd. (step 2), were observed under a microscope, and henceforth, images were captured and laborious and carefully ground-truth marked under expert supervision at Guwahati Neurological Research Center (GNRC) (step 3). Image acquisition was followed by pre-processing (step 4 and 5). Subsequently, five different texture features (step 6) were extracted and were compared using five standard classifiers (step 7). The outcomes are given in the results section.

Data Collection

The tissue blocks from the Neurology Department of GMCH. GMCH is the foremost Public Sector Hospital in the region catering to the general public and treating all diseases. Hematoxylin and Eosin (H&E) staining were done at Ayursundra Healthcare Pvt. Limited. Ayursundra Healthcare Pvt. Ltd. is one of the biggest diagnostics providers of the region with a strong pathology unit. On staining the block with H&E, the nucleus is marked blue and cytoplasm as pink for better understanding. After the staining is done, we can study the features of the slide under the microscope. The microscopic image capture was done at Guwahati Neurological Research Centre (GNRC). Established in 1987, GNRC is the first private super-specialty healthcare center in North East India. Childhood medulloblastoma is a grave case of the tumour since the most commonly affected age-group of children is a vulnerable age, and the survival rate of the disease is also almost nil. There is no online benchmark dataset of biopsy images available for different classes of this tumor. **Fig. 2** also illustrates the difference in the microscopic appearance[12] of each subtype. The classic pattern (b) has a sheets-like arrangement of the cells with extensive cellularity. The large cell (c) has the same structural arrangement as the classic subtype. However, the size of the cells of large cell medulloblastoma are enlarged and, unlike the classic variant, is highly aggressive. The nodular pattern (a) has nodules like structure in the distribution of cell arrangement. Finally, the last subtype of desmoplastic (d) has a network of collagen fiber.

The study design was cross-sectional, wherein data (a small quantity of the tissue block) were collected from children (of age<15 years) who were diagnosed with this type of tumor. The samples were collected as a part of the post-operative procedure by the collaborating medical institutions, namely the Neurology Department of GMCH. The diagnosis of the same sample of the tissue block manually done in GMCH was noted down, for corroboration with our findings.

During the study period, a total of 15 children were treated from whom the samples were collected. A graphical presentation of the flow of participants is shown in **Fig. 3**.

We obtained the images at 10x microscopic low power view of resolution 2048px×1536px. It is tough to get images corresponding to each subtype. We collected 4 classics, 2 desmoplastic, 1 nodular and 2 large cell anaplastic cases/blocks from which we captured a total of 44, 38, 23 and 28 images respectively. However, to have an unbiased classifying accuracy, we finally selected 23 best images of each subgroup for our purpose, with the help of experts. A link to a few images acquired for the study is provided in *https://figshare.com/s/44220013b8ba5c2f075f.*

Image Preprocessing

The difference in the expertise of the technician preparing the histological slides landed us with images that were either dark or lightly stained. Hence preprocessing methods were carried out such that variation in staining does not affect our result. For this color, channeling was first done to extract the most appropriate channel to get our region of interest(ROI). Color channeling is a prime task for image processing, as individual color channels provide distinct and diverse information about the image. It is an integral part, as not all color information is required. The pathway that carries the targeted information can be considered neglecting the other available channels. Hence, for selecting the appropriate channel for our case, the image was converted to all available color channels viz. RGB, HSV, YCbCr, CMY, and L*a*b* shown in **Fig. 4,** and then the particular color channel in which the foreground was differentiated best from the background was chosen.

The choice of the color channel is purely based on visual perception. Also, we are only interested in the cells and not cytoplasm of the cells, so we followed the perception that since

the cells are purple after undergoing H&E staining and purple is closer to red, so red channel is visually most precise. Secondly, image enhancement was performed to differentiate the targeted ROI from the background. Next, we analyzed all enhancement methodologies available, as depicted on a selected image in. **Fig. 5.**

The application of the filters to our images demonstrated that none gave us the required outcome that we were searching for, namely the best elimination of the background while focusing on the cell alignment of the tissue samples. Hence, we customized a filtering mask by filtering the original image by the customized filter kernel [1,1,1; 1,10,1; 111] of size 3×3. The enhanced images were grayscale (the second image of row 1 of Fig. 5). The obtained images were brighter, but it highlighted the cell alignment while giving a clean background, which was needed.

Feature Extraction

Next, experiments were performed for recognizing the different subtypes of childhood medulloblastoma. We were mainly concerned in the cell distribution, as different subtypes have a different spatial distribution of cells. Commonly, shape and color are the significant features used by the pathologist for analyzing histological images. Shape-based characteristics generally define the physical structure of an object like area, perimeter, orientation, etc.[17]. Now both these features are observed through 40x level microscopic view. However, in our case, the 10x microscopic view is concerned with only the spread of cells in the tissue region and does not carry any shape information. The architectural difference is a visual perimeter observed by a pathologist based on human intelligence. We experiment if this human intelligence can be included in digital histopathology using computer vision by texture features. In general, the texture is the surface characteristic and description of an object, stated by its size, density,

arrangement, and proportion of elementary parts. For this experiment, we studied the effect of both first order and second order textural features of the tissue samples. The first-order statistics define individual pixel statistical properties like mean, variance, skewness, and kurtosis. The second-order statistics represent the relation between pixels[17]. We considered five well-established texture features, namely Gray level co-occurrence matrix (GLCM)[18], Gray level run length matrix (GRLM)[19], Local binary pattern (LBP), First-order histogram feature[20] (HOGL) and Tamura[21].

GLCM: GLCM describes how frequently ~~a~~ couple of pixels appears together in the image via spatial association among pixels of the image. It signifies the likelihood of arriving from pixel *i* to *j* in a definite direction. We applied this well know texture feature in all four directions, usually represented by degrees (0^o, 45^o, 90^o, 135^o). The four directions represent the neighborhood of the pixel *i*. Each of the elements (*i,j*) in the GLCM matrix gives the occurrence of the pixel with value '*i*' in the directions to the pixel with value '*j*.' The degrees considered represent the offset between a pair of pixels. For this study the offset was measured by a single-pixel distance taken as [0 1; -1 1; -1 0; -1 -1] for all four directions. We extracted the features of pixel identity pairs relating to contrast, correlation, energy, and homogeneity. The mathematical formulae of the features are as follows:

Every element (*i,j*) in GLCM specifies the number of times that the pixel with value occurred adjacent to a pixel with value *j* in image P.

$$\text{Contrast} = \sum_i^j |i - \text{j}|^2 \text{p(i; j)} \quad (1)$$

It returns the contrast between the pixel and its neighborhood in the entire image. Contrast property is also said to be the variance and inertia of the image. For a constant image, the value of contrast is 0.

Correlation $=\sum_{i,j}(i-\mu i)(-\mu j)p(i,j)/\sigma i\sigma j$ (2)

μ and σ represent the mean and standard deviation of pixel *i* and pixel *j*.

It returns a measure of how perfectly two neighbors are correlated. The correlation of an image is always in the range of [-1,1]. The association for an entirely positive correlated image is +1, and the perfect negatively correlated image is -1.

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

The range of Energy is between [0,1]. It returns the sum of squared elements in the GLCM. The energy is also referred to as uniformity or angular second moment of the image. For a constant image, the energy is 1.

Homogeneity $=\sum_{i,j}P(i,j)/(1+|i-j|)$ (4)

Homogeneity of an image returns a value which evaluates the proximity of the distribution of elements in the GLCM with respect to the GLCM diagonal. It ranges from [0,1]. From here we extracted sixteen (4× 4) GLCM texture features.

GRLM: Secondly, we used GRLM feature extraction. GLCM[22] describes the presence of pixel pairs in a particular direction whereas GRLM denotes the associated path of particular pixels in a definite direction[22]. The gray level run is a line of pixels in a certain direction having the same intensity value[23]. Gray level run length is the number of such pixels and the number of occurrences of such is the run-length value. Here, we considered the run length as the number of neighboring pixels that possess the same grey intensity in horizontal to its right direction. Since the resolution of the images is large the images are quantized to 16 gray levels for ease of computational complexity and speedy performance. The features that were extracted using GRLM were short-run emphasis (sre), long-run emphasis (lre), gray level non-uniformity (gln),

run percentage (rp), run length non-uniformity (rln), low gray level run emphasis (lgre) and high gray level run emphasis (hgre). From here we extracted seven textural features. The mathematical formulae[24] are given below.

$$sre = 1/n\sum_{i,j} p(i, j)/j^2 \quad (5)$$

$$lre = 1/n\sum_{i,j} j^2 p(i, j) \quad (6)$$

$$gln = 1/n\sum_{i}(\sum_{j} p(i, j))^2 \quad (7)$$

$$rp = \sum_{i,j} n/p(i, j)j \quad (8)$$

$$rln = 1/n \sum_{j}(\sum_{i} p(i, j))^2 \quad (9)$$

$$lgre = 1/n\sum_{i,j} p(i, j)/i^2 \quad (10)$$

$$hgre = 1/n\sum_{i,j} i^2 p(i, j) \quad (11)$$

Here $p(i, j)$ is the number of times there is a run of length j having gray level i, n is the total number of pixels.

Tamura Texture Feature: Our third feature set is the Tamura[21] based feature set. The Tamura feature computes six texture features viz coarseness, contrast, directionality, line-likeness, regularity, and roughness. The authors[25] reported that the first three are more significant than the rest. Therefore, we extracted the primary three considerable features of coarseness, contrast, and directionality from Tamura based features. The coarseness signifies the number of prominent spatial dissimilarity of grey levels, which are the small texels constituting the textures of the image. Directionality takes into account the edge contour and the direction angle. These are formulated using 'prewitts' edge detector operator.

HOGL: Our fourth feature set was the histogram-based feature, which was mean, variance, skewness, kurtosis, energy, entropy, and probability of intensity distribution of the grayscale image. The mean, variance, skewness, kurtosis is said to be the main first-order texture statistics to represent an image[17]. The variance examines how an individual pixel value deviates from the mean value of the pixels in the image. Skewness contributes to whether an appeared surface is darker or brighter than the average. Kurtosis studies the peak of the probability distribution. A high value of kurtosis will have a sharper peak. From here, we extracted 14 textural features.

LBP: For LBP, a binary code is produced for each pixel by thresholding the neighborhood with the value of the center pixel. The neighborhood considered was 8 pixels, which entail all the pixels surrounding the center pixel with a range or offset value of 1 considering the immediate neighbor of the center pixel. We did not encode any rotation features and found the mapping as rotation invariant. A histogram is then produced for different binary codes generated representing various types of spots, curved edges, at areas, etc in the image textons[17]. The LBP features encode local texture information and can be used for many tasks including classification, detection, and recognition. Lahdenoja[26] referred LBP to be associated with statistical and structural texture analysis. It was used to check the local image contrast in the subtypes. From here, we extracted 59 textural features.

Hence in total, 99 textural features were extracted from the above mentioned five pure feature sets and analyzed for classification.

Classification

The classification task involves five significant steps. *a)* Create the necessary test and training dataset. *b)* Find the best- fit algorithm to classify the data. *c)* Train the model. *d)* Predict the subtype using the fitted model and finally *e)* to attempt to increase accuracy by selecting the

optimum feature set. We implemented five standard classifiers, viz. Linear Discriminant (LD), Quadratic Discriminant (QD), Support Vector Machine (SVM), k Nearest Neighbor (KNN), and Decision Tree-based classification methodologies[27-33] to classify all the images using pure texture feature sets *i)* individually and also *ii)* in combinations of individual features. Both LD and QD were used since we do not know whether we had the same or different covariance for the classes. The SVM classifier was computed using the 'quadratic' kernel and 'one vs. all' classification model. For the KNN classifier 'Euclidean distance' was used as the distance measure with 10 number of nearest neighbors. The decision tree model used 'Ginni index' as the splitting criteria. We used these five different classifiers since, in machine learning, there is no particular rule as to which classifier will perform best for a given feature set. These classifiers are few of the prominent machine learning classifiers[34] and were chosen based on their diverse nature of classification. The merit and demerit of each classifier are presented in **Table. 1**. The combined features sets were based on varied concatenations of the pure sets as follows: most accurate 2 (best-2), most accurate 3 (best-3), most accurate 4 (best-4) or all five pure feature sets (all-5). The accuracy of the samples is calculated as Accuracy = (True positive + True negative) = Number of samples/100.

Construction of training and testing data set

We have used 5 fold cross-validation where the data set was divided into 5 subsets, and the holdout method is repeated 5 times. Hence for each iteration, 80% of the data are treated as training samples, and the remaining 20% of testing samples. For each iteration, one of 5 subsets was used as a test set, and 5-1=4 set was used as the training set. The advantage of this method is that every data point gets to be in the training and test set, removing any biases associated with the data point role in training or testing.

RESULTS

Performance evaluation

Table:2 shows the contribution of different feature sets towards the multiclass classification of the various subgroups in childhood medulloblastoma. First, we tested the accuracy of LBP, GLCM, HOGL, Tamura, and GRLM of individual feature sets and then arranged the feature sets in descending order of accuracy of classification[9]. Successively, we checked the classification accuracy for the combinations of feature sets. The results are then grouped by best-2 (LBP+GRLN), best-3 (LBP+GRLM+GLCM), best-4 (LBP+GRLM+GLCM+Tamura) and all-5 (LBP+GLCM+GRLM+Tamura+HOGL) feature sets, as shown in **Table. 2**.

Fig. 6(a) depicts the classification accuracy of feature sets, along with the performance time (in seconds) shown in **Fig. 6(b)**. Correlation analysis was performed to confirm the efficiency of feature combinations. The best-4 feature set formed by the combination of GLCM, GRLN, LBP, and Tamura features gave us the best performance for classification in terms of accuracy.

Although accuracy is a good indicator of achievement, it is effectual only when datasets are symmetric, i.e., where values of false positive and false negatives are almost the same. Therefore, to further validate the performance of SVM, we calculated a few other measures viz: Precision, Recall, False positive rate, Specificity, and Kappa. The values of the tests are shown in **Table. 3**.

As revealed, SVM has a better value for the other measures as too. Therefore, we plotted a Receiver operational characteristics (ROC) curve shown in (**Fig. 7**) to analyze the area under the curve (AUC) for each subclass.

Class 3 had a perfect AUC=1 followed by class 4, class 1, and class 2. We did a correlation analysis for this set of the concatenated 83-dimensional feature vector. We found that there is little correlation between individual feature subsets, which reveal that the concatenated feature set measures different aspects of features of the image efficiently, and combining them might be useful in the extraction of information. **Fig. 8** shows the correlation analysis of the best-4 feature set.

The confusion matrix (**Fig. 9**) pictorially describes the accuracy of predicted and actual class for the best-4 feature set. The diagonal elements give the number of correctly classified class instances. It is seen that for class 1 and 2 we got a small misclassification where an instance of class 1 is classified as class 2 while in class 2 some cases are classified as of class 1 and class 4. Classes 3 and 4 are correctly classified. The numbers 1 to 4 are the four subgroups of childhood medulloblastoma.

LD and QD presented the least accurate performance since they assume the feature set to be independent. In contrast, in a real situation, it is difficult to obtain independent sets and indeed our observed features were dependent. Moreover, it requires the distribution to be normal, but in practice, our distribution may not follow the normal distribution. SVM gave the best performance since SVM performs well for non-linear separable classes, and our subtypes are not linearly separable. Decision Tree did not perform well and might have got stuck into local minima. KNN was a good performer as our feature sets had less irrelevant samples, as proved during correlation analysis.

Feature Reduction

As explained above, our proposed approach gives the highest accuracy for the combined best-4 feature i.e., combining GLCM, GRLM, Tamura, and LBP feature set and excluding the HOGL based

feature set. Also, it is observed that for classification SVM classifier gives the highest accuracy for this combination. This model corresponding to features from GLCM, GRLM, LBP and Tamura contains 83 features. Our next target was to reduce this feature set to reduce computational time further. We used the well-applied feature reduction methodology of principal component analysis (PCA) for dimensionality reduction. PCA is used for feature reduction by reducing the number of correlated variables in a feature set by using Eigenvalues and Eigenvectors. The SVM classifier was used for PCA since it performed best in accuracy for the mentioned feature set. We calculated the accuracy of the classifier based on the consideration of the number of principal components. First, we calculated the accuracy at an interval of 10, starting with 10 principal components. On reaching 30, we found the accuracy decreases from 96.7% (at 20 principal components) to 94.6%. The accuracy decreased as we increased the number of components, as clear from **Fig. 10**.

To find out the exact number of components for utmost accuracy, we performed our experiment taking a variable number of principal components lesser than 30. Fig. 10 includes the table showing the number of components vs. accuracy, which reveals that we can achieve the best accuracy of 96.7% using 20 principal components. We thus reduced the feature set for the combination of best-4 feature set from 83 to 20 and increased accuracy by 5.4%.

DISCUSSION

The individual texture features were evaluated using five different classifiers. The SVM and KNN perform better in the case of all the feature sets. The LBP feature set gives the highest classification accuracy of 84.8%, and the HOGL based feature set provides the least certainty of 69.6% using the SVM classifier. The accuracy of classifiers for the five classifiers in decreasing order are LBP, GRLM, GLCM, Tamura, and HOGL. Different texture sets measure various aspects

of texture. Hence to explore which aspect provided improved results, we repeated the classification procedure in a combination of feature sets. For this purpose, we combined best-2, best-3, best-4 and all-5 feature set based on their performance measured in terms of accuracy. We got an efficiency of 91.3% by using best-4 texture feature set followed by best-2, best-3, and all-5 feature set. Significantly, it is observed that the combination of all five feature set reduces the classification accuracy. We performed a feature reduction for our best-fit feature set and found that the efficiency increases as we decrease the feature set from 83 to 20.

Observations

Difficulties

The architectural property of medulloblastoma subtypes are very minute and requires experience and skill to evaluate such abnormalities. As follow-up treatment depends highly on its type, it is vital to have an accurate diagnosis. The time taken for analysis of the biopsy sample in the region requires considerable time. There is a very less ratio of the pathologist to patients. A fact which is right for all underdeveloped and developing parts of a country. As prognosis is very sensitive, most of the cases move outside the region for better prospects. This results in the lesser availability of data in the area. The lack of a benchmark data set makes it further challenging to carry out such research. But since its classification can give a longer survival rate, we have tried to attempt to classify its subgroups using computer-aided diagnosis.

Comparison

Till now, multiclass textural classification for medulloblastoma is not reported in any study. Further, as mentioned, there is no benchmark dataset available where our methodology could be applied and compared with our indigenous dataset. Hence, a comparison of our results was not possible. The multi-class analysis is more complicated than a two-way classifier problem.

Researchers have distinguished medulloblastoma into two types, anaplastic and non-anaplastic. Galaro et al[6] have classified anaplastic and non-anaplastic medulloblastoma using a bag of words classifier and obtained an accuracy of 87%. Rao et al[7] had used CNN based classification for anaplastic and non-anaplastic medulloblastoma for 10 images and obtained an accuracy of 89.8%. Lai et al[8] classified between anaplastic and non-anaplastic and using Haar, Haralick, and Law textural feature and achieved an AUC of 91%. Our proposed method for multiclass classification obtains a higher accuracy of 96.7%.

CONCLUSION

The paper presents a multiclass classification among childhood medulloblastoma sub-groups using five textural feature sets. We evaluated the accuracy of classifiers using independent feature set and in the groups of best-2, best-3, best-4, and all-5 feature set. Using best-4 feature set, we were able to achieve an accuracy of 91.3%, which further increases to 96.7% on feature reduction using principal component analysis. It is also seen that not all combination of feature sets gives us a better result. Since individual feature sets provide different textural information, it is recommended that we use the group of features for better classification. As observed, existing methods are well applicable for pre and post-processing of the images, giving satisfactory accuracy. This accuracy will increase once our dataset increases, and we may be able to use deep learning methodologies for the system.

ACKNOWLEDGEMENT

We would like to convey our sincere thanks to Dr Basanta Kr. Baishya, Head of the Department of Neurosurgery, Guwahati Medical College, and Dr Inamul Haque, MCH Trainee from the Department of Neurosurgery, Guwahati Medical College for providing us with the tissue blocks and Dr Anup Das from Ayursundra Healthcare Pvt. Ltd. for processing the slides. We would

further thank Dr Shabnam Ahmed of Guwahati Neurological Research Centre, Dispur for dedicating her time and effort and helping us in image acquisition and providing the ground truth. We are grateful to the Institute of Advanced Study in Science and Technology (IASST), Guwahati, for giving us the platform to perform our research.

AUTHORS' CONTRIBUTIONS

D.D. had done the practical implementation of the work. Dr L.B.M. supervised the whole idea of the work and was a conceptual support team member. Dr S.A. was the medical expert in the study and analysis of our work. Dr B.K.B. and Dr I.H. were from the neurosurgery department who provided us with the samples and also shared knowledge regarding them.

CONFLICT OF INTEREST

NO, I declare that the authors have no competing interests or other interests that might be perceived to influence the results and/or discussion reported in this paper.

DISCLOSURE STATEMENT

The author(s) declare no competing interest, either financial or non- financial.

REFERENCES

1. Polednak, A. P.& Flannery, J. T. (1995). Brain, other central nervous system, and eye cancer. *Cancer*, 75, 330-337.
2. Muzio, B. D. & Gaillard, A.F. (2016). WHO classification of CNS tumours: Retrieved from https://radiopaedia.org/articles/who-classification-of-cns-tumours-1.
3. Kool, M., Korshunov, A., Remke, M., Jones, D. T., Schlanstein, M., Northcott, P. A., Cho, Y. J., Koster, J., Schouten-van & Meeteren, A., (2012). Molecular subgroups of medulloblastoma: an international meta-analysis of transcriptome, genetic aberrations, and clinical data of wnt, shh, group 3, and group 4 medulloblastomas. *Acta Neuropathol*. 123, 473-84.

4. Davis, F. G, Freels, S., Grutsch, J., Barlas, S. & Brem, S. (1998). Survival rates in patients with primary malignant brain tumors stratified by patient age and tumor histologic type: an analysis based on surveillance, epidemiology and end results (seer) data. *Neurosurg*. 88,1-10.
5. Chhabda, S., Carney, O., DArco, F., Jacques, T. S. & Mankad, K. (2016). The 2016 world health organization classification of tumours of the central nervous system: what the paediatric neuro-radiologist needs to know. *Quant Imaging Med Surg*, 6(5), 486-489.
6. Galaro, J., Judkins, A. R., Ellison, D., Baccon, J. & Madabhushi, A. (2011). An integrated texton and bag of words classifier for identifying anaplastic medulloblastomas. *In: Proc. Annu. Int. Conf. IEEE Eng. Med. Biol. Soc. p. 34433446.*
7. Roa, A. C., Arvalo, J., Judkins, A., Madabhushi, A. & Gonzlez, F. (2015). A method for medulloblastoma tumor differentiation based on convolutional neural networks and transfer learning. *In: Proc. 11th International Symposium on Medical Information Processing and Analysis*.; vol. 9681.
8. Lai, Y., Viswanatha, S., Baccon, J., Ellison, D., Judkins, A. R. & Madabhushi, A. (2011). A texture-based classifier to discriminate anaplastic from non-anaplastic medulloblastoma. *In: Proc. IEEE 37th Annu. Northeast Bioeng. Conf.*
9. Kather, N. J, Cleo-AronWeis, Bianconi, M., Melcher, R. L., S & Schad, M., Timo Gaiser, Marx, A., Frank Gerrit Zllner. (2016). Multi-class texture analysis in colorectal cancer histology. *Scientific Reports*. 6.
10. Awwad, A., Rodriguez, D., Jaspan, T., Grundy, R & Auer, P. D. (2012). T1 based texture analysis for the classification of paediatric medulloblastoma, ependymoma and their anaplastic subtypes. *In: Proc. 36th European Society of Neuroradiology (ESNR) Annual Meeting, Edinburgh.*
11. Manual, U. 2017. Sparse textures. Retrieved from https://docs.unity3d.com/Manual/SparseTextures.html.
12. Graham, I. D. &Lantos, L.P. (2015). Greenfields neuropathology. (ninth ed).; vol. 2. CRC Press, USA.
13. Uppuluri, A. (2008). GLCM texture features.; Retrieved from https://in.mathworks.com/matlabcentral/fileexchange/22187-glcm-texture-features
14. Wei, X. (2007). Histogram features of a gray level image: Retrieved fromhttps://in.mathworks.com/matlabcentral/fileexchange/17537-histogram-features-of-a-gray-level-image.
15. Sornapaudi, S. (2016). Tamura features. Retrieved from https://github.com/Sdhir/TamuraFeatures.
16. Karunanithi, R. (2010). Gray level run length matrix. Retrieved from [https://in.mathworks.com/matlabcentral/fileexchange/26694-gray-level-run-length-matrix].
17. Sree Sharmila, T & Ramar, K. (2014). Texture feature extraction.; Retrieved from http://shodhganga.inibnet.ac.in/bitstream/10603/24460/9/09_chapter4.pdf.
18. Galloway, M. M. (1975). Texture analysis using gray level run lengths. *Computer Graphics Image Process*, 4, 172-179.

19. Tang, X. (1998). Texture information in run-length matrices. *In: Proc. IEEE Transactions on Image Processing.*
20. Malik, F. & Baharudin, B. (2013). The Statistical Quantized Histogram Texture Features Analysis for Image Retrieval Based on Median and Laplacian Filters in the DCT Domain. *The International Arab Journal of Information Technology,* 10(6), https://iajit.org/PDF/vol.10,no.6/5112.pdf
21. Tamura, H., Hideyuki, Mori, S. &Yamawaki, T. (1978). Textural features corresponding to visual perception. *Systems, Man and Cybernetics*, 6, 460-473.
22. Singh, R. K. (2016). A comparison of grey-level run length matrix and grey-level co-occurrence matrix towards cereal grain classification. *International Journal of Computer Engineering and Technology*. 7(6), 9-17.
23. Hafsha, W.N, Kairuddin, W. & Hafizah, W.M. (2017). IOP Conf. Ser.: Mater. Sci. Eng. 226 012136
24. Manavalan, R. S. & Thangavel, K. (2012). Evaluation of textural feature extraction from GRLM for prostate cancer TRUS medical images. International Journal of Advances in Engineering and Technology, 2, 401-409.
25. Patel, M. J. & Gamit, C. N. (2016). A review on feature extraction techniques in content based image retrieval. *In: Proc. IEEE WiSPNET conference.*
26. Lahdenoja, O. (2005). Local binary pattern feature vector extraction with CNN. *In: Proc. 9th International Workshop on Cellular Neural Networks and Their Applications.*
27. Luque, A., Romero-Lemos, J., Carrasco, A. & Barbancho, J. (2018). Non-sequential automatic classification of anuran sounds for the estimation of climate-change indicators. *Expert Systems with Applications*, 95,248-260.
28. Härdle, W. K., & Simar, L. (2012). Applied multivariate statistical analysis. (4th Ed.). Springer Science & Business Media.
29. Huang, T. M. & Kecman, V. (2005). Gene extraction for cancer diagnosis by support vector machines – an improvement. *Artificial Intelligence in Medicine*, 35,185–194.
30. Muthu Rama Krishnan, M., Chakraborty, C., Paul, R.R. & Ray, A.K. (2012). A Hybrid segmentation, characterization and classification of basal cell nuclei from histopathological images of normal oral mucosa and submucous fibrosis. *Expert Systems with Applications*, 39, 1062–1077.
31. Mandal, M. K. (2014). High voltage electrode system design by application of support vector machine. Thesis submitted to Jadavpur University, Kolkata, India. Available from: http://docplayer.net/11598278-High-voltage-electrode-system-design-byapplication-of-supportvector-machine-manas-kumar-mandal.html. Accessed 27 October
32. Kumar, R., Srivastava, R. & Srivastava, S. (2015). Detection and Classification of Cancer from Microscopic Biopsy Images Using Clinically Significant and Biologically Interpretable Features. Journal of Medical Engineering, Volume 2015, 14 pages. doi:10.1155/2015/457906
33. Cover, T. M., & Hart, P. E. (1967). Nearest neighbor pattern classification. IEEE Transactions on Information Theory, 13 (1), 21–27.
34. Dharmadhikari, C. S., Ingle, M. & Kulkarni, P. (2011). Empirical studies on machine learning based text classification algorithms. *Advanced Computing: An International Journal,* 2(6),161-169.

Figure Captions:

Fig. 1. Block Diagram for our proposed work.

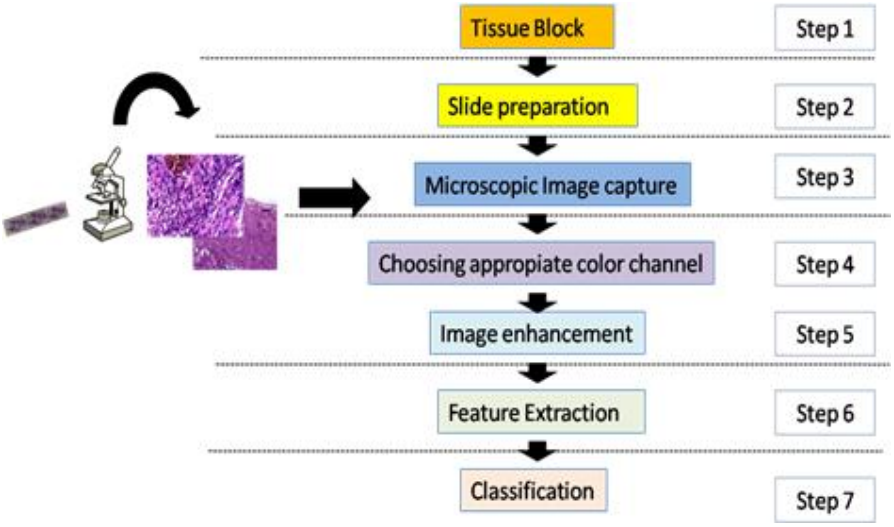

Fig. 2. Figure showing the microscopic view of our data samples a) nodular b) classic c) large cell d) desmoplastic subtype.

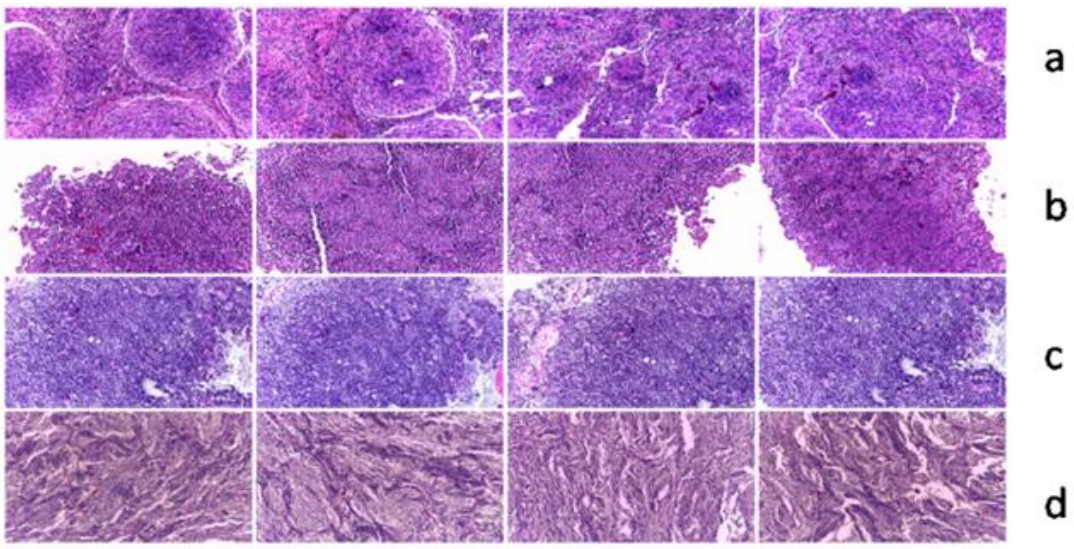

Fig. 3. Data Collection procedure.

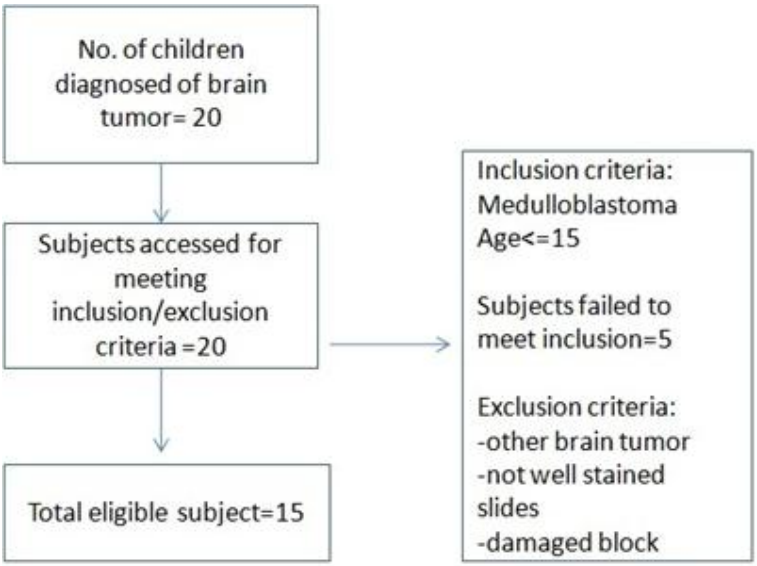

Fig. 4. Showing the different color channels for our dataset images.

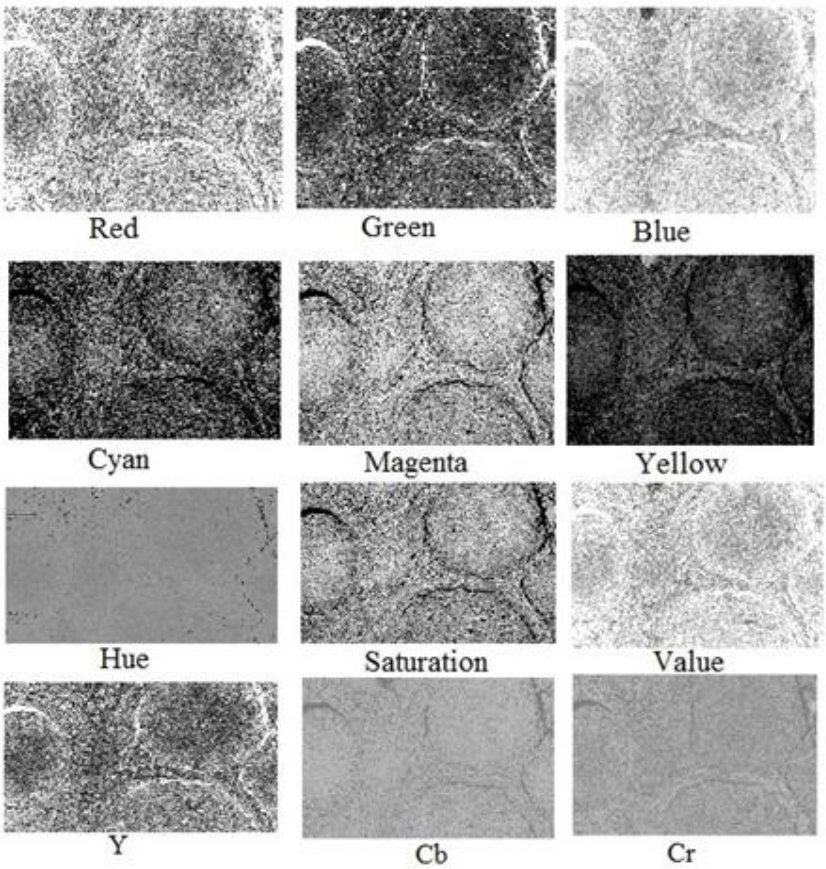

Fig. 5. Result of different enhancement methods applied to our images.

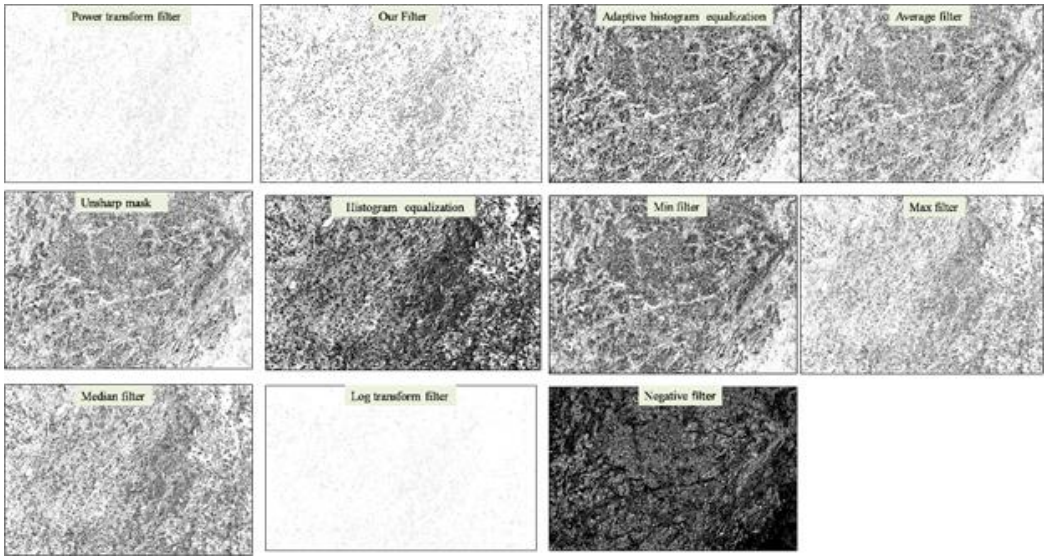

Fig. 6(a). Plot of Classification Accuracy of the used subgroup of features with different classifiers.

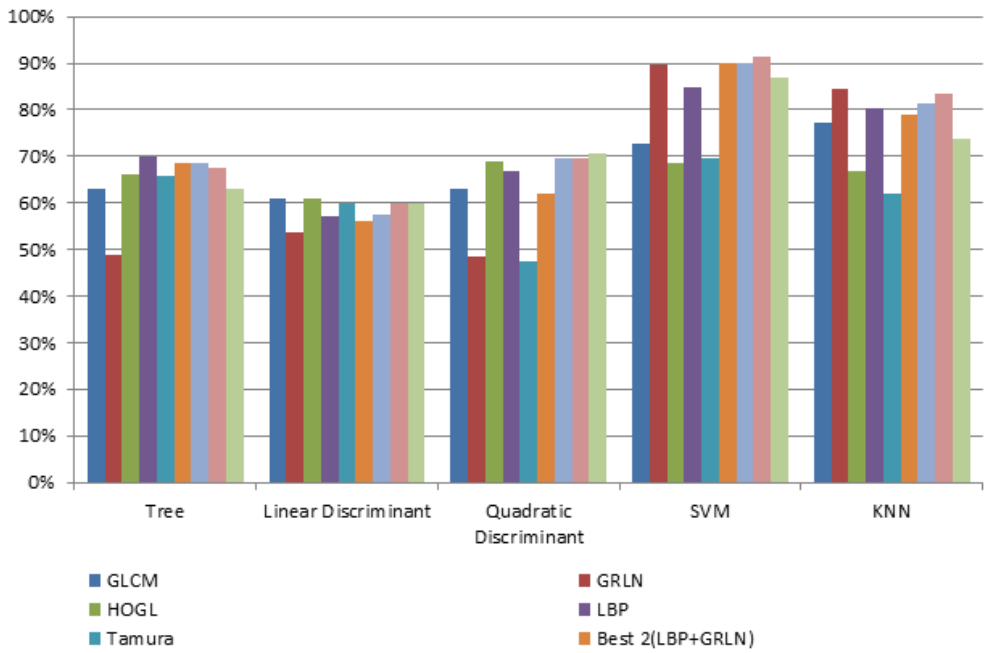

Fig. 6(b). Training time analysis of the feature sets for SVM classifier.

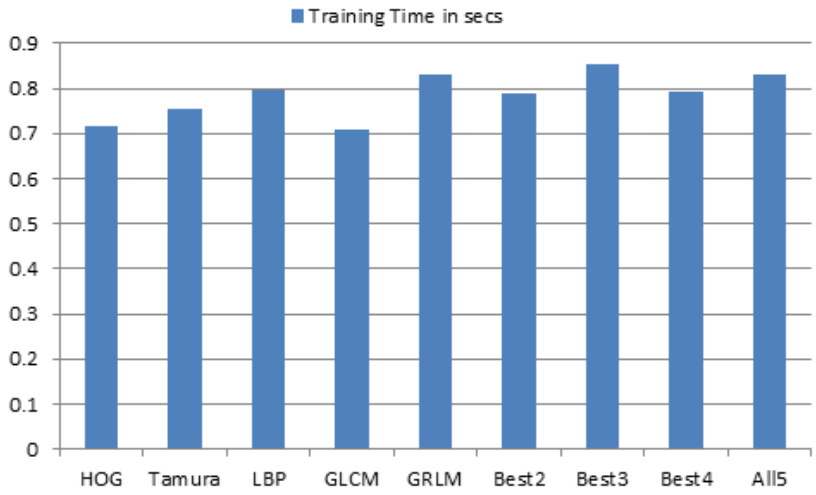

Fig. 7. ROC curve for Best4 feature set for the different classes.

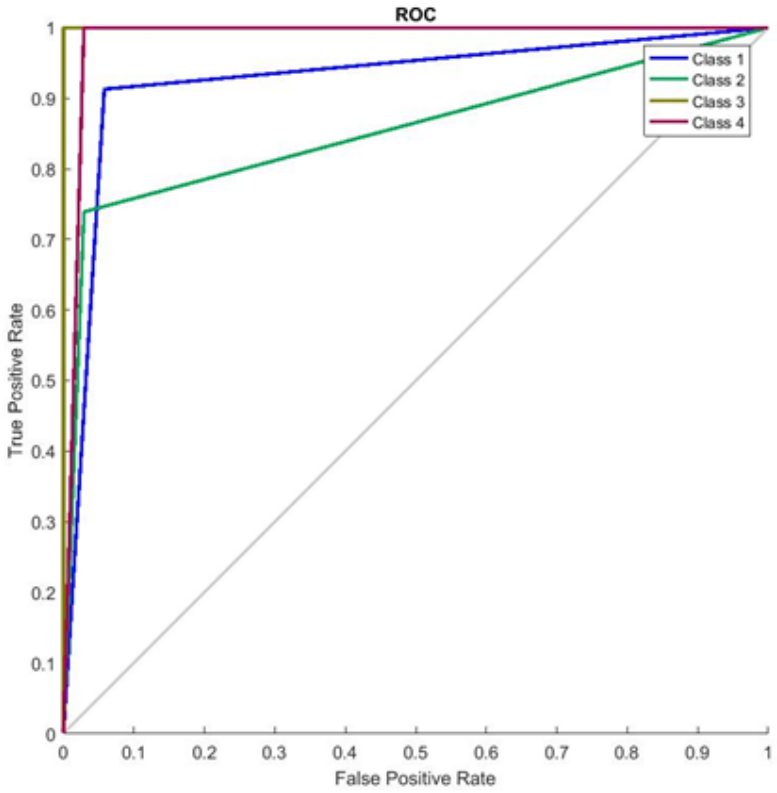

Fig. 8. Correlation analysis of the features.

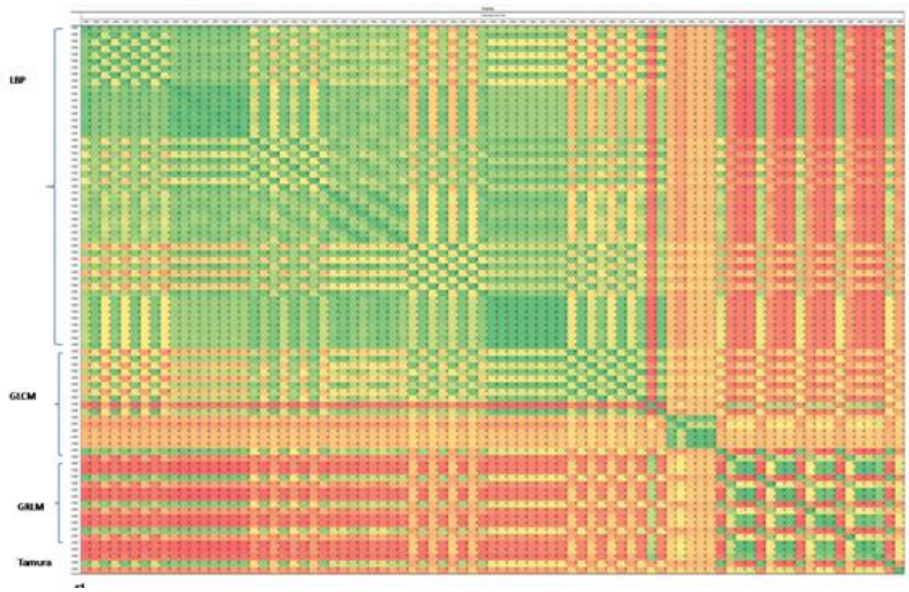

Fig. 9. Confusion matrix showing the true class and predicted class output of best-4 feature set.

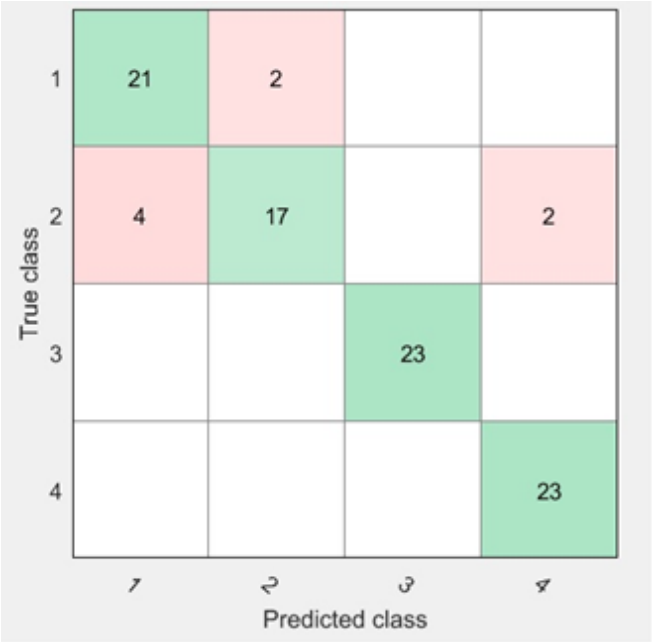

Fig. 10. PCA analysis showing accuracy vs no. of components curve of best-4 feature set.

Number of principal components	Accuracy
5	90.2%
10	93.5%
13	92.4%
14	93.5%
15	91.3%
16	94.6%
17	95.7%
18	94.6%
19	93.5%
20	**96.7%**
21	96.7%
22	95.7%
23	94.6%
25	94.6%
30	94.6%
40	88%
50	75%
60	53%
70	30.4%
80	25%

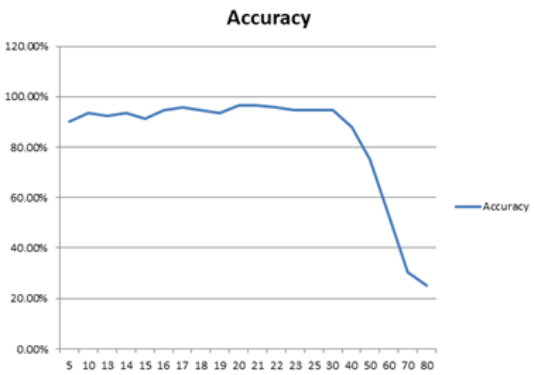

Table. 1. Comparison of different classifiers used.

Classifiers	Advantage	Disadvantage
KNN	It is efficient, simple, non-parametric and is easily implemented	It has a slow performance rate when the size of the training sample increases. The performance is hindered in the presence of irrelevant samples.
LD/QD	Requires normal distribution. The requirement of training samples is less for parameter estimation in classification. Large speed and accuracy are achieved when applied to a large dataset.	The algorithm works well with independent features only. It gives a poor performance for dependant feature set.
Decision Tree	It can be used for any datatype. Non-parametric. It does not have any distribution requirement. It is faster for the larger feature set.	The major risk in the implementation of a decision tree is that with the presence of an alternative tree. It can get stuck at local minima.
SVM	No distribution requirement. It is able to manage large spaces of the feature. Good for nonlinear feature	It has high usage and computational time

Table. 2. Performance comparison in % of individual and best-2, best-3, best-4 and all-5 texture.

Individual features	Classifiers				
	Tree	LD	QD	SVM	KNN
GLCM	63	60.9	63	72.8	77.2
GRLM	48.9	53.6	48.5	89.7	84.5
HOGL	66	60.9	69	68.5	67
LBP	70	57	67	84.8	80.4
Tamura	65.7	59.8	47.4	69.6	62
Best2(LBP+GRLM)	68.5	56	62	90.2	79
Best3(LBP+GRLM+GLCM)	68.5	57.6	69.6	89.9	81.5
Best4(LBP+GRLM+GLCM+Tamura)	67.4	59.8	69.6	91.3	83.6
All5(LBP+GRLM+GLCM+Tamura+HOGL)	63	59.8	70.7	87	73.9

Table. 3. The performance measure for the best-4 feature set.

Classifiers	Precision	Recall	False Positive Rate	Specificity	Accuracy	Kappa
SVM	.913	.913	.03	.97	.913	.884
Tree	.67	.67	.09	.90	.674	.565
QD	.69	.68	.06	.93	.696	.594
LD	.59	.513	.11	.89	.598	.464
KNN	.83	.87	.05	.95	.836	.78