Classification of Intracranial Hemorrhage CT images for Stroke Analysis with Transformed and Image-based GLCM Features

Santwana Gudadhe (santwana.gudadhe@pccoepune.org)
Pimpri Chinchwad College of Engineering https://orcid.org/0000-0002-9894-854X

Anuradha Thakare
Pimpri Chinchwad College of Engineering

Research

Keywords: Intracranial Hemorrhage, Computed tomography, Random Forest, Random Tree, REPTree, Discrete cosine transform, Classification, Gray level co-occurrence matrix, Discrete wavelet transform.

Posted Date: January 17th, 2022

DOI: https://doi.org/10.21203/rs.3.rs-1234293/v1

Classification of Intracranial Hemorrhage CT images for Stroke Analysis with Transformed and Image-based GLCM Features

Santwana S. Gudadhe[1], Anuradha D. Thakare[2]

[1] Computer Engineering Department, PCCOE, Pune, Savitribai Phule Pune University
santwana.gudadhe@pccoepune.org

[2] Computer Engineering Departments, PCCOE, Pune, Savitribai Phule Pune University
Adthakare2014@gmail.com

* Corresponding Author: Santwana S. Gudadhe,
santwana.gudadhe@pccoepune.org

Abstract –

Intracranial hemorrhage (ICH) is one of the severe types of brain stroke. Artery burst results in bleeding inside the brain and its surrounding tissue. Depending upon brain bleed location, hemorrhage gets classified. This paper presents a hybrid feature selection approach to form joint feature vector sets using transformed image features and image-based gray level co-occurrence matrix (GLCM) texture features. Feature extraction is performed by applying discrete wavelet transform, discrete cosine transform, and GLCM features. Feature vector selection model is built with the combination of Discrete wavelet transform, Discrete cosine transform, and GLCM features. A joint feature vector is formed with transformed image features and GLCM features. These joint feature vectors form an acute input for the further classification process. The machine learning algorithm i.e. Random Tree, Random Forest, and REPTree are used for the classification of Intracranial hemorrhage CT images. The classification results obtained are further analyzed for accuracy considering transformed features and GLCM features. It is observed that Random forest classifier results in the highest classification accuracy of 87.97% for discrete wavelet transform and GLCM (DWT+GLCM) feature set

Keywords: Intracranial Hemorrhage, Computed tomography, Random Forest, Random Tree, REPTree, Discrete cosine transform, Classification, Gray level co-occurrence matrix, Discrete wavelet transform.

1. Introduction

Brain hemorrhage, a life-changing event for patients. Accurate diagnosis plays an important role in the medical field to improve the patient's condition in a faster way. Intracranial hemorrhage (ICH) is a heterogeneous disease. It includes intracerebral hemorrhage and subarachnoid hemorrhage [1]. Intracerebral hemorrhage is caused due to bleeding inside the brain tissue itself. whereas subarachnoid hemorrhage is caused due to bleeding in the surrounding brain space. ICH diagnosis mainly involves a patient physical examination and medical history. Accurate and fast diagnosis is crucial as 50% of patients' death happen within 24 hours 35 to 52 % of patients enters in critical zone and die within a month as mortality reaches up to 60% after 30 days [2].To perform accurate diagnosis, first stage examination is performed by a physician using non-contrast computed tomography (CT) analysis. CT image analysis helps to identify and locate the bleeding location inside the brain [3].

Stroke is an emergency condition of the brain which requires immediate attention. To deal with, it involves major challenges: accurate ICH diagnosis, time-consuming process related to decision making, sometimes insufficient experience in the case of neophyte radiologists, and may involve lack of experience in the decision-making process. Traditional methods deal with a patient's physical examination and CT image analysis. Based on expert radiologist knowledge treatment will be given. To assist the radiologist, there is a need to provide an automated diagnostic tool that helps the radiologist to fasten the decision-making process to provide faster treatment to the patients.

Recently, research studies have mainly focused on the Classification, identification, and detection of intracranial hemorrhage with the aim of diagnosis time reduction and fastening the clinical workflow. Specifically, the classification system helps doctors perform correct analysis of the images and be able to classify between hemorrhage and normal images. This paper presents the concept of a hybrid feature selection approach to form joint feature vectors using transformed image features and image-based gray level co-occurrence matrix (GLCM) texture features. Transformed feature vector generation using discrete cosine transform and discrete wavelet transform along with GLCM based feature vector. Finally, classification accuracy is calculated for Random forest, Random tree, and REPTree machine learning classifier.
This article aims to provide:

- Intracranial hemorrhage CT image feature extraction using transform-based image features and GLCM based texture image feature approach.
- Performance evaluation of Machine Learning classification algorithms for the hybrid feature selection approach.

The rest of this paper is organized as follows. Literature survey discussed in Section 2. Material and Methods are discussed in Section 3. The hybrid feature selection approach is detailed discussed in subsection 3.2. Result analysis & discussion are presented in Sections 4 and 5. Finally, Section 6 discussed the conclusion.

2. Literature survey

In recent studies, Image analysis is the key area in the medical field. Image analysis helps radiologists to perform diagnoses more accurately. In the case of Brain stroke analysis identification, classification, and prediction of hemorrhage is gaining more attention in recent years. Accurate diagnosis of Intracranial hemorrhage plays a crucial role in patients' recovery. In the image classification process, feature extraction is an important part.

Accurate and optimal extracted feature sets can speed up the classification task. Texture feature selection is one of the most promising methods used for accurate diagnosis. Researchers use the gray level co-occurrence matrix-based, first-order features, transform approach with machine learning classifier to correctly classify the query image. Paper [4] uses the shape features, First-order, and gray level co-occurrence matrix-based texture features from ROI image to classify the hemorrhage vs normal image using Random Forest classifier. [5] Presented Brain Haemorrhage CT scan type classification with Multi-Layer Perceptron algorithms and K-Nearest Neighbour using Gray Level Co-occurrence Matrix-based Feature extraction method. [6] presented a gray level run length feature extraction method based on the wavelet approach. Support Vector Machine (SVM) classifier is applied to obtain segmented images and classify the brain soft tissues from CT ICH images. Feature optimization has been performed by using the MDEE algorithm and BDM.

Feature extraction from regions of interest to minimize the processing time helps to speed up the execution and diagnosis time. Many researchers targeted ROI-based feature extraction to perform classification. Discriminative features extraction from a region of interest (ROI) brain CT images and classification using Support Vector Machines, Neural Networks, Decision Trees classifiers have been experimented with [7]. Identification of the regions of ICH on Diffusion-Weighted (DW) brain images [8] using cosine transform and wavelet transform has experimented. Feature analysis and classification are performed using the K-Nearest Neighbor (KNN) classifier. Best performance achieved for the wavelet transform.

2.1 Image-based Feature extraction

Feature extraction based on image texture is one of the promising methods. Grey Level Co-occurrence Matrices (GLCM) is the feature extraction method based on image texture characteristics.[9]. It uses the concept of the pixel intensity distribution. It is mainly used for X-ray analysis, which contains black, white, and different shades of gray.

In the image for every pixel, the homogeneity value is calculated and if any change is observed then there is the highest change of different textures. In the brain, the abnormal area has a different texture from the rest of the part of the area. If there is a drastic change in the matrix value, then there is a maximum chance of getting an abnormal region[9]. This paper uses Grey Level Co-occurrence Matrices features as an Image-based feature. extraction Calculation of selected features listed below [10]:

- Contrast - It is a measure of intensity contrast change of a pixel and its neighboring pixel.

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

- Energy - It calculates the total number of square elements in the GLCM matrix. The values are in between 0 and 1.

$$Energy = \sum_{i,j=0}^{N-1} P_j(ij)^2 \quad (2)$$

- Homogeneity - it is a calculation of smoothness of gray level distributed in the image.

$$Homogeneity = \sum_{i,j=0}^{N-1} \frac{P_{i,j}}{(1+(i-j)2} \quad (3)$$

- Correlation - It is a measure of gray level correlation from the pixel to the neighboring pixel.

$$Correlation = \sum_{i,j=0}^{N-1} P_{i,j}\left(\frac{(i-\mu)(j-\mu)}{\sigma^2}\right) \quad (4)$$

2.2 Discrete Cosine Transform (DCT) based Feature Extraction-

DCT transformed-based feature extraction involves the generation of transformed images from input images by applying cosine transform.

Once the cosine transforms are applied to the input image, high energy coefficients are assigned to low energy areas. Transform images include the high energy and low energy coefficients. These high energy coefficients are used to form feature vectors by taking the standard deviation and mean of coefficients. 2-D DCT image of size MxN is given below [11,12].

$$F(x,y) = \alpha(x)\alpha(y)\sum_{u=0}^{N-1}\sum_{v=0}^{M-1} cos\left[\frac{\pi x}{2N}(2u+1)\right]cos\left[\frac{\pi y}{2M}(2v+1)\right]f(u,v) \quad (5)$$

$$\alpha(x)\alpha(y) = \left\{\sqrt{\frac{1}{N}}\, for\ x,\ y \neq 0\ ;\ \sqrt{\frac{2}{N}}\, for\ x,\ v = 0\right\} \quad (6)$$

where f(u,v) is pixel intensity at point (u,v). MxN is the size of an image.

2.3 Discrete Wavelet Transform (DWT) based Feature Extraction -

Discrete wavelet transform uses high-pass and low-pass filters to filter the input image. Discrete Wavelet transforms suppress the noise. Haar transform is the simplest transform considered for the discrete wavelet transform. Because of its simplicity Haar transform is taken as a reference for the wavelet transform. Once the 2-D discrete wavelet transform (DWT) is applied to the input image it decomposes the input image into different frequency bands with the use of high pass filter and low pass filters successively [13] and the input image returns the approximation coefficient matrix along with detail coefficient matrices. Mean and standard deviation are evaluated from the approximate coefficient matrix and detail coefficient matrix. These values are considered as DWT-based transformed-based feature vectors. The DWT is defined as [14].

$$W_{\varphi}\left(j_0, k\right) = \frac{1}{\sqrt{M}} \sum_{x} f(x) \varphi_{j0,k}(x) \tag{7}$$

$$W_{\psi}(j, k) = \frac{1}{\sqrt{M}} \sum_{k} f(x) \psi_{j,k}(x) \tag{8}$$

where f(x).$\varphi_{j0,k}(x)$ and $\psi_{j,k}(x)$ are functions of discrete variables x=1,2….M-1.

3. Material and Methods

3.1. Materials

Standard Intracranial Hemorrhage CT image dataset for Brain Stroke is collected from Kaggle [15]. The dataset includes a total of 2501 CT images including Hemorrhage and Normal images. The dataset includes 30 Hemorrhage patient data and 50 normal patient data. For each patient on average, 30 CT image slices are available.

The experimentation is performed on the MATLAB R2020b platform. Feature extraction is performed from the Intracranial Hemorrhage CT image dataset. Extracted feature vectors are maintained and passed as an input to machine learning classifiers. Figure 1 represents a sample image for the Intracranial Hemorrhage CT image dataset for Brain Stroke. For experimentation a total of 1825 CT images are considered (1000 Normal images and 825 Abnormal images). All images are in .jpg file format. The original images are of 640 x 640 size. For the processing, it has been reduced to 256 x 256 size. Data Preprocessing is done by applying a Gaussian filter to the images as it consists of noises.

Using The Gray Level Co-occurrence Matrix (GLCM) texture features and Image transformed features are considered. From the Machine Learning WEKA tool, Random Forest, Random Tree, and REPTree Classifiers are used to classify the normal and abnormal images. Finally, Accuracy for these classification algorithms is calculated and analyzed.[14].

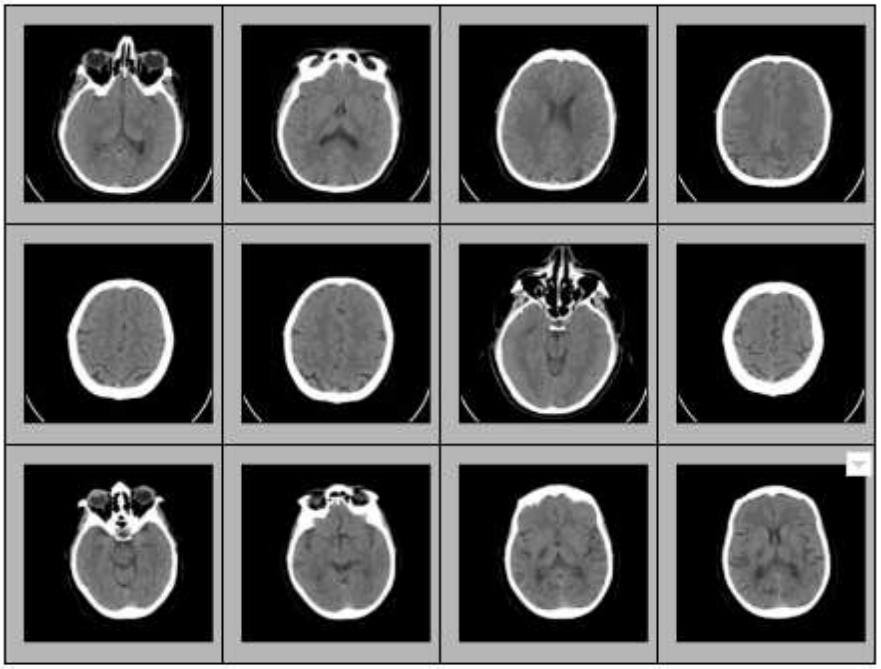

Fig.1 Shows a few sets of images from Dataset.

3.2. Method

The proposed Intracranial Hemorrhage classification system includes the concept of transformed image features and texture features. Proposed work includes CT image preprocessing, transformed feature generation and selection, classification, and comparative analysis of various feature selection methods for Random Forest classifier, Random Tree Classifier, and REPTree Classifier. Figure 2a & 2b shows the Proposed hybrid Feature extraction Method and Classification model.

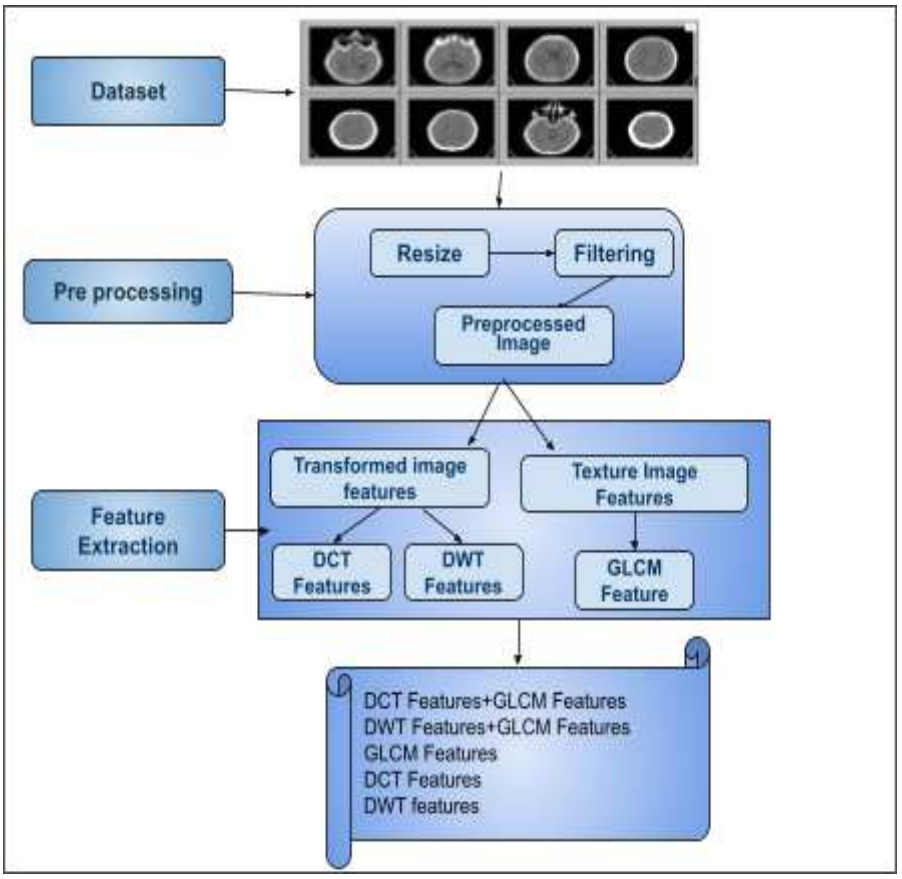

Fig. 2 (a) Proposed Hybrid Feature Extraction method.

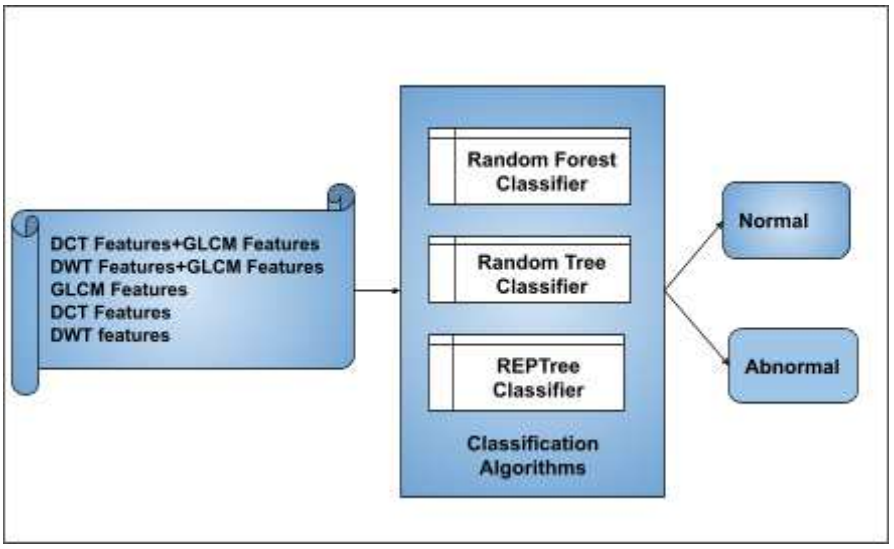

Fig.2 (b) Shows the Classification model.

3.2.1 Proposed hybrid Feature Extraction Method for Intracranial Hemorrhage classification

The proposed feature extraction method includes the joint feature set from transformed-based features and GLCM based features. Feature extraction using a GLCM based approach helps to identify the texture of an image and it is a key idea to classify the images from one texture to another. Also, transformed-based approaches help to identify and segregate the high frequency and low-frequency coefficients of the input images. So, images containing important information can easily be obtained from the energy coefficients. By using both the approach, to accurately classify the normal and abnormal images and to speed the classification task, these two approaches are combined to generate different hybrid feature vector sets and classification accuracy analyzed for Random forest, Random tree, and REPTree classifier. Transformed image-based features, GLCM based features, and hybrid approaches for feature selection are discussed in this section along with Random forest, Random tree, and REPTree classifiers are used to classify the images.

A. Transformed image features generated for Intracranial Hemorrhage classification

Transformed image features are generated using Discrete Cosine transform Coefficient and Discrete wavelet transform Coefficient. The first 2-D Discrete cosine transform is applied on a preprocessed input image and high-frequency coefficient and low-frequency coefficient matrix are obtained. These DCT coefficients are further considered for feature vector generation. To form the DCT transformed feature vector, DCT transformed images are further processed. Mean and Standard deviations of a transformed image are considered as DCT transformed features. Similarly, Discrete wavelet transform is also applied on preprocessed input images, and DWT coefficients are further considered for feature vector generation. To form the DWT transformed feature vector, DWT transformed images are further processed. Mean and Standard deviations of a transformed image are considered as DWT transformed features.

B. Image-based Features generated for Intracranial Hemorrhage classification

Grey Level Co-occurrence Matrices (GLCM) method is used for Image-based Features. Considered GLCM features are contrast, Correction, energy, and homogeneity.

C. Hybrid feature selection approach for Intracranial Hemorrhage classification

This section discusses a hybrid approach for feature selection which includes features based on energy coefficient and texture features.
To form a group of feature vector DCT transformed feature coefficients, DWT transformed feature coefficient and GLCM texture features are used. And finally, these Feature vectors are passed as an input to the Classification Algorithm to classify the normal and abnormal images. Table 1 depicts the hybrid approach for feature selection. The first hybrid approach considered is Discrete cosine transform features and Image-based GLCM features.

The second approach combines the discrete wavelet transform and Image-based GLCM features along with Images based Features, DCT based features, and DWT-based features.Hybrid feature selection Algorithm discussed below.

Algorithm : Hybrid feature selection approach (HFS)

1: **procedure** HFS (M, s, t, CF,V) ◁ M is the dataset of s Intracranial Hemorrhage CT images. t= number of features. CF= set of classifiers used.

Extract Hybrid feature set (s)

i. Apply gaussian filter on s

ii. Obtained a Transformed feature vector using DCT & DWT transform.

iii. Apply discrete cosine transform on filter image.

iv. Obtained high frequency and low-frequency coefficients from the images

v. Generate transformed feature vectors by taking the mean and standard deviation of the image frequency coefficients.

vi. Obtained Image-based GLCM feature vector i.e contrast, Correction, energy, and homogeneity

vii. Apply Hybrid feature selection approach {Transformed FV+ GLCM FV)

viii. Apply the classifier

ix. return t

end procedure

2: **Procedure** Test Classifier (CF, Ac) CF is a classifier. The testing will be k-fold cross-validation.

return c with best accuracy measure Ac

end procedure

Table 1: Hybrid approach for feature selection.

Feature selection methods	Feature Considered
DCT+GLCM	DCT Transformed Mean, Standard Deviation, contrast, Correction, energy, homogeneity, and Entropy
DWT+GLCM	DWT Transformed Mean, Standard Deviation, Contrast, Correction, Energy, homogeneity, and Entropy
DCT	DCT Transformed Mean, Standard Deviation
DWT	DWT Transformed Mean, Standard Deviation
GLCM	Contrast, Correction, Energy and homogeneity, and Entropy

3.2.2. Machine Learning Classifiers

To classify the Intracranial hemorrhage CT image several machine learning classifiers were tested. Classification accuracy is measured for different algorithms like Naive Bayes, Decision tree, Random forest, Random tree, and REPTree classifier. Especially those that performed well have been discussed below. The objective of this work is to correctly classify brain CT images as abnormal and normal images.

Random Forest

Random forest Classifier is an ensemble-based supervised machine learning model [16]. During the training phase, it includes the construction of multiple decision trees from dataset Features. Random Forest is the most popular prediction model. It performs the

final prediction by taking the average of each predicted decision tree. Random tree classifiers can handle the missing data.

Random Tree

The random decision tree is a supervised machine learning model. It includes multiple decision tree generation from different samples of training data. In image classification, each pixel gets classified by calling the decision tree. Rank order is calculated based on the number of decisions made. Random tree [17] performs classifications with every tree and outputs the class label that received the majority of votes.

REPTree

REPTree [18] is also known as a fast decision tree classification algorithm. Entropy and information gain of the features along with reduced error pruning is used to build the decision tree. The back fitting method is used to perform pruning of REPTree. REPTree makes the multiple trees in different iterations by using the logic of the regression tree.

4 . Results Analysis

Classification accuracy for the proposed hybrid feature selection approach along with Image features and transformed-based features are calculated using machine learning classifiers like the random forest, random tree, REPTree, Naive Bayes, and decision tree classifier. After performing analysis, final classification results for the random forest, random tree, and REPTree have been presented over here. results obtained for DCT transformed Feature extraction based Classification, DWT transformed Feature extraction based Classification, Image Features (GLCM) based Classification, DCT transformed Features and Image Feature-based Classification, and DWT transformed Features and Image Feature-based Classification has been discussed in this section.

4.1. DCT transformed Feature extraction based Classification

For the proposed system, classification accuracy is calculated for DCT transformed image features with different k-fold values are 5,10,15,20,25, and 30. Random Tree, Random Forest, and REPTree machine learning classifiers are used. Table 2 depicts the classification accuracy for different k values with Random Tree, Random Forest, and REPTree classifiers. In all classification algorithms, Random forest shows a better result as compared to Random Tree and REPTree classifiers. The random forest classifier for k=25 shows the maximum classification accuracy is 75.12%. The Random Tree classifier for k=10 shows the maximum classification accuracy is 68.10% and the REPTree classifier for k=20 shows the maximum classification accuracy is 65.80%. Figure 3 shows the comparative analysis of classification accuracy of Random Tree, Random Forest, and REPTree machine learning classifiers for different k values.

Table 2: DCT Transformed Feature-based Classification classification accuracy with Random Tree, Random Forest, and REPTree classifiers

Classifier	**Classification Accuracy(%) for K- fold values**					
	k=5	**k=10**	**k=15**	**k=20**	**k=25**	**k=30**
Random Forest	72.16	73.91,	75.06	74.4,	75.12	74.57
Random Tree	66.52,	68.10,	64.27	65.15	65.04	68.10
REPTree	63.50,	64.38,	65.64,	65.80	65.75	64.98

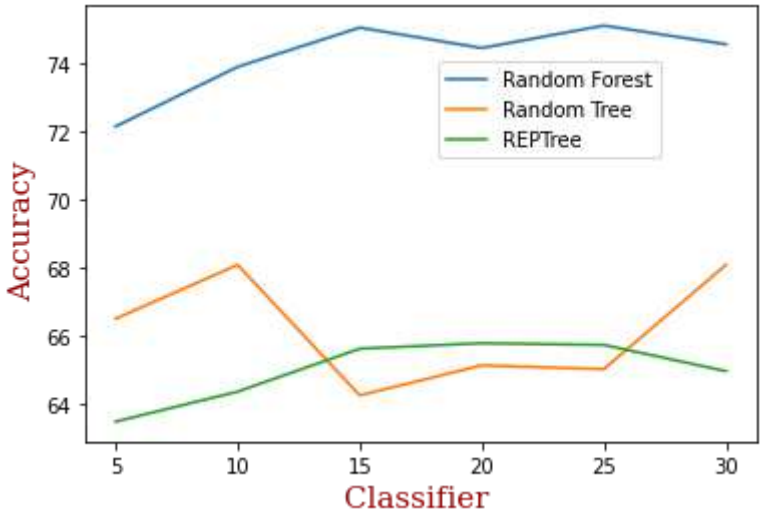

Fig. 3 Comparative analysis of classification accuracy of Random Tree, Random Forest, and REPTree with DCT based features

4.2 DWT Transformed Feature-based Classification

The classification accuracy calculated for DWT transformed image features with different k-fold values is 5,10,15,20,25 and 30. Random Tree, Random Forest, and REPTree machine learning classifiers are used. Table 3 depicts the classification accuracy for different k values with Random Tree, Random Forest, and REPTree classifiers. In all classification algorithms, Random forest shows a better result as compared to Random Tree and REPTree classifiers. The random forest classifier for k=20 shows the maximum classification accuracy is 82.35%. The Random Tree classifier for k=15 shows the maximum classification accuracy is 75.39% and the REPTree classifier for k=25 shows the maximum classification accuracy is 72.16%.

Figure 4 represents the comparative analysis of classification accuracy of Random Tree, Random Forest, and REPTree machine learning classifiers for different k values.

Table 3: DWT Transformed Feature based Classification classification accuracy with Random Tree, Random Forest, and REPTree classifiers

Classifier	Classification Accuracy(%) for K- fold values					
	k=5	k=10	k=15	k=20	k=25	k=30
Random Forest	80.60	81.47	81.42	82.35	82.30	82.13
Random Tree	72.10,	74.95,	75.39,	73.75	74.30	74.46
REPTree	69.80,	71.39,	71.83,	72.10	72.16	71.45

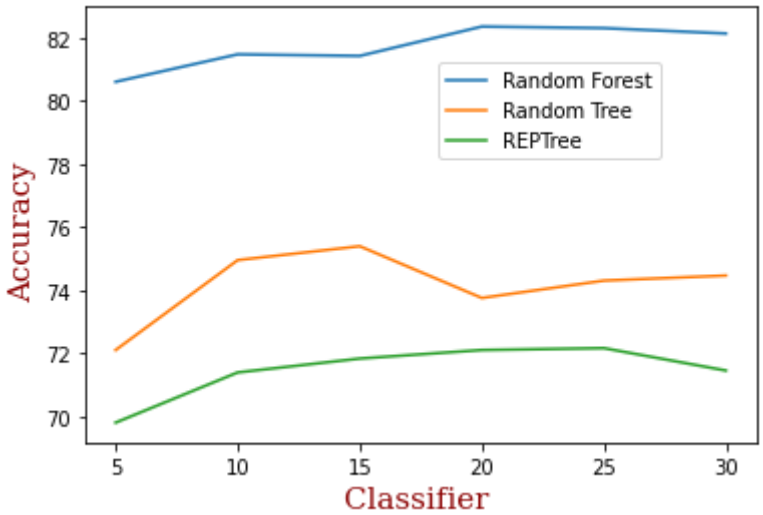

Fig. 4 Shows the comparative analysis of classification accuracy of Random Tree, Random Forest, and REPTree with DWT features

4.3 Image Features based Classification

For the proposed system, Image first-order features and Gray-level co-occurrence matrix (GLCM) feature-based classification accuracy is calculated for Random Tree, Random Forest, and REPTree machine learning classifiers with different k-fold values are 5,10,15,20,25 and 30. Table 4 depicts the classification accuracy for different k values with Random Tree, Random Forest, and REPTree classifiers. The random forest classifier for k=20 shows the maximum classification accuracy is 86.90%. The Random Tree classifier for k=10 shows the maximum classification accuracy is 79.89% and the REPTree classifier for k=25 shows the maximum classification accuracy is 77.26%. Figure 5 represents the comparative analysis of classification accuracy of Random Tree, Random Forest, and REPTree machine learning classifiers for different k values.

Table 4: Image-based GLCM Feature set classification accuracy for different k values with Random Tree, Random Forest, and REPTree classifiers

Classifier	**Classification Accuracy(%) for K- fold values**					
	k=5	k=10	k=15	k=20	k=25	k=30
Random Forest	75.23,	85.97,	86.13,	86.90	86.35	86.41
Random Tree	78.08,	79.89,	79.56,	79.72	79.01	79.12
REPTree	75.23,	75.72,	74.13,	76.16	77.26	76.87

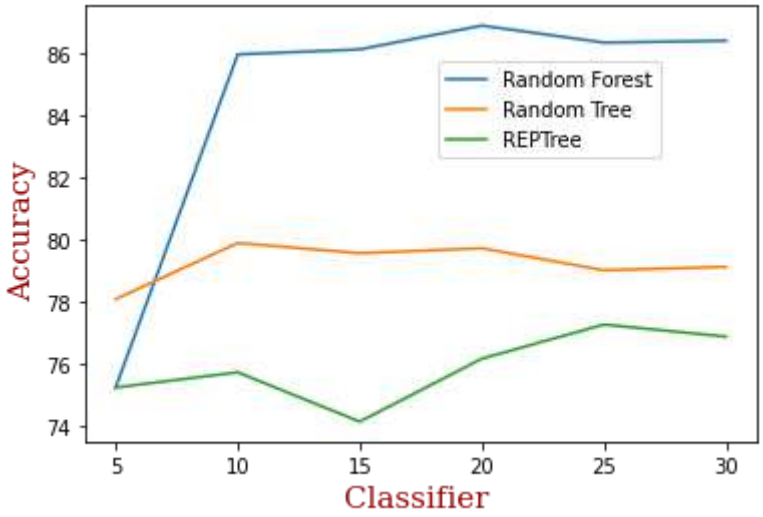

Fig. 5 Shows the comparative analysis of classification accuracy of Random Tree, Random Forest, and REPTree with Image-based features

4.5 DCT transformed Features and Image Feature-based Classification

Classification accuracy for Random Tree, Random Forest, REPTree is calculated using DCT transformed feature and Gray-level co-occurrence matrix (GLCM) image features. For evaluation different k-fold values considered are 5,10,15,20,25 and 30. Table 5 depicts the classification accuracy for different k values with Random Tree, Random Forest, and REPTree classifiers. Random forest classifier for k=20 & k=25 shows the maximum classification accuracy is 87.67%. The Random Tree classifier for k=10 shows the maximum classification accuracy is 80.49% and the REPTree classifier for k=25 shows the maximum classification accuracy is 78.84 %. Figure 6 represents the comparative analysis of classification accuracy of Random Tree, Random Forest, and REPTree machine learning classifiers for different k values.

Table 5: DCT transformed Features and Image Feature-based Classification accuracy for different k values with Random Tree, Random Forest, and REPTree classifiers

Classifier	Classification Accuracy(%) for K- fold values					
	k=5	k=10	k=15	k=20	k=25	k=30
Random Forest	85.97,	86.95,	87.28,	87.67	87.67	86.95
Random Tree	78.68,	80.49,	78.24,	79.45	79.50	78.84
REPTree	74.90,	76.82,	74.41,	77.36	78.84	75.50

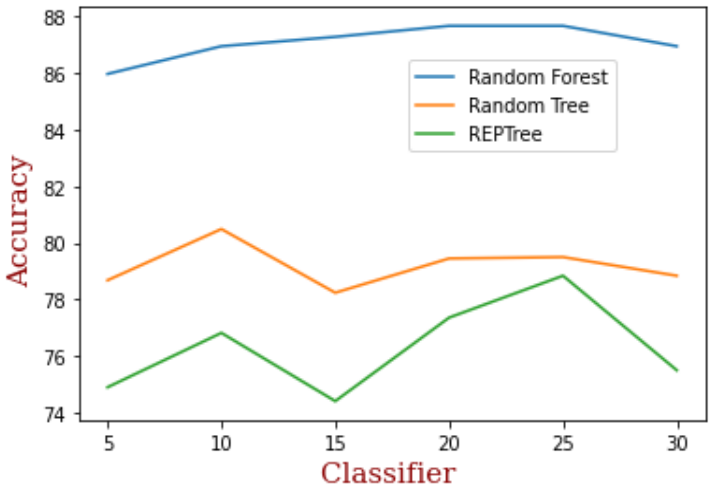

Fig. 6 Shows the comparative analysis of classification accuracy of Random Tree, Random Forest, and REPTree with DCT transformed based s and Image Feature-based features.

4.6 DWT transformed Features and Image Feature-based Classification.

Classification accuracy is calculated using DWT transformed features and Gray-level co-occurrence matrix (GLCM) image features for Random Forest, Random Tree, and REPTree classifiers with different values of k = 5,10,15,20,25 & 30. Table 6 depicts the classification accuracy for different k values with Random Tree, Random Forest, and REPTree classifiers.

Random forest classifier for k=20 & k=25 shows the maximum classification accuracy is 87.67%. The Random Tree classifier for k=10 shows the maximum classification accuracy is 80.49% and the REPTree classifier for k=25 shows the maximum classification accuracy is 78.84 %. Figure 7 shows the comparative analysis of classification accuracy of Random Tree, Random Forest, and REPTree machine learning classifiers for different k values.

Table 6: DCT transformed Features and Image Feature-based classification accuracy for different k values with Random Tree, Random Forest, and REPTree classifiers.

Classifier	**Classification Accuracy(%) for K- fold values**					
	k=5	k=10	k=15	k=20	k=25	k=30
Random Forest	86.52,	87.39,	87.39,	87.89	87.72	87.67
Random Tree	78.84,	80.32,	80.49,	80.21	79.50	81.20
REPTree	77.86	78.02	75.89	78.52	78.35	77.58

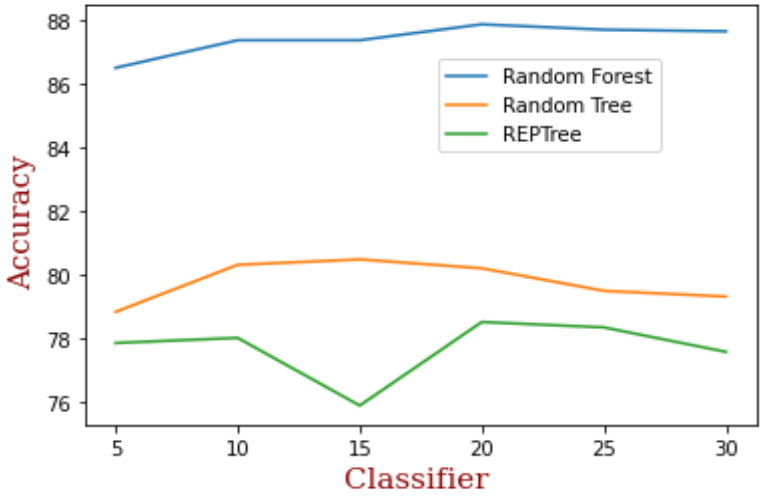

Fig. 7 Shows the comparative analysis of classification accuracy of Random Tree, Random Forest, and REPTree with DWT transformed-based and Image Feature-based features.

5. Discussion

For the proposed Intracranial Hemorrhage classification system, transformed-based features using DCT and DWT transform and GLCM based features are considered. Classification accuracy for Random Forest, Random Tree, and REPTree classifiers is calculated considering the different values of k = 5, 10,15,20,25 & 30. In all classification models, Random forest shows the better result in terms of accuracy. DCT transformed image feature group-based classification shows a satisfactory result. DWT transformed-based feature group shows better results over DCT transformed feature-based classification. From the results, it is observed, Random forest shows the overall improved performance for k=20 and 25 for various feature groups. In the case of the Random Tree, classifier performance is better for the k=10 value and REPTRee shows better performance for the k=25 value.

Considering image-based feature groups, classification accuracy shows a better value for all classifiers as compared to DCT transform feature and DWT transformed feature group.

From the experimental results, it is observed that considering combined DWT transform features and GLCM features (**DWT+GLCM)** group, classification accuracy for the random forest is most prominent as compared to other feature groups. Similarly for DCT transform features and GLCM features (**DCT+GLCM)** group results show better classification accuracy. Table 7 depicts the comparative analysis of Random Forest classification accuracy for feature groups.

Figure 8 and 9 represents the comparison between various feature group classification accuracy for Random Forest Classifier with k=20. .

Table 7: Comparative analysis of Random Forest classification accuracy for Hybrid feature groups along with Transformed based and Image-based features.

Feature group	**Random Forest Classification Accuracy.(%) for k=20**
DCT+GLCM	87.67
DWT+GLCM	**87.89**
DCT	74.46
DWT	82.35
Image Features	86.90

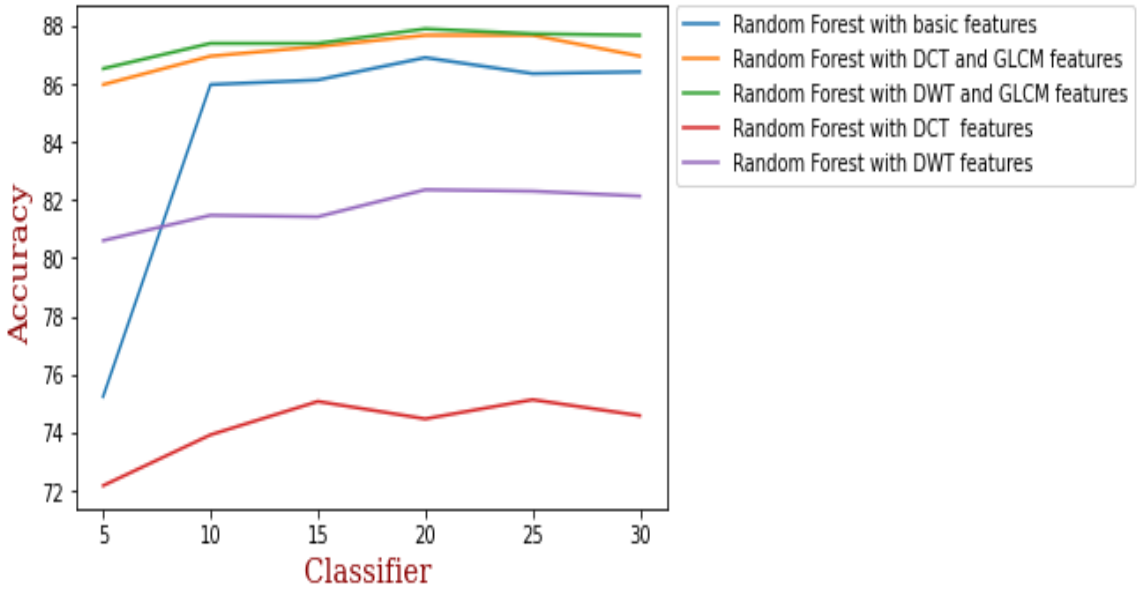

Fig. 8 Shows the comparative analysis of Random Forest classification accuracy with different Feature sets.

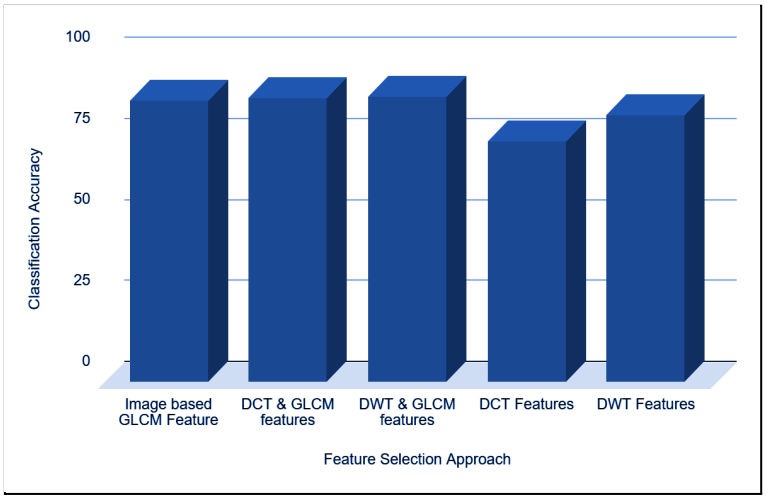

Fig. 9 Shows the comparative analysis of Random Forest classification accuracy with different Feature sets.

Intracranial hemorrhage CT images. Figure 10 represents the ROC curve analysis for the Random forest classifier. Considering Normal and abnormal images, The Intracranial hemorrhage CT images were considered true positive and CT images without Intracranial hemorrhage were considered as a true negative. CT image without Intracranial hemorrhage, but models considered it as an abnormal image are false-positive and Intracranial hemorrhage image falsely interpreted as normal image considered as a false negative. False-positive rate and True positive rate were plotted to generate the ROC curve.

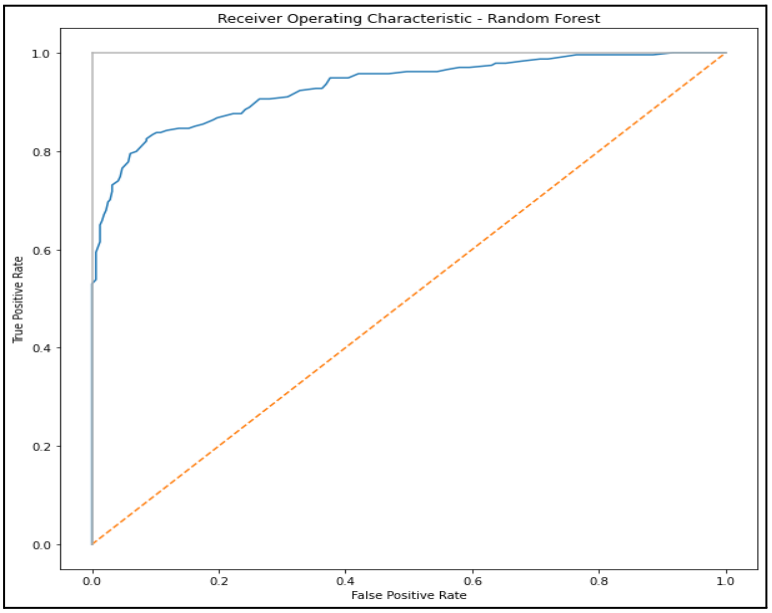

Fig. 10 ROC curve analysis for Random forest classifier

6. Conclusion

In this paper, intracranial hemorrhage CT images classification has been performed using Random Forest classifier, Random Tree Classifier, and REPTree classifier. A hybrid feature-based approach is used with the combination of image-based gray level co-occurrence matrix features and transformed-based features. The Combined feature-based approach helps to classify the intracranial

hemorrhages CT images in a more accurate way as compared to the image-feature-based and transformed feature-based approach. High energy coefficients obtained from CT images by applying the transformed approach are the key finding in the transformed-based feature approach. Discrete Wavelet transform provides a better image compression as compared to Discrete Cosine Transform results in optimal features. Considering the hybrid feature-based classification accuracy for DCT+GLCM and DWT+GLCM is analyzed. DWT + GLCM hybrid feature group shows the maximum classification accuracy is 87.89 for random forest classifier with k=20, considering DWT + GLCM hybrid feature group. followed by, DCT + GLCM able to achieve maximum classification accuracy is 87.67 with k =20. For the result analysis, the hybrid features-based approach shows better performance as compared to image-based GLCM features and transformed feature-based approach. In all Classification algorithms, Random Forest shows improved accuracy. In the future, hybrid feature-based approaches can be extended with more detailed feature groups for large datasets.

Abbreviations

CT : Computed Tomography

ICH : Intracranial Hemorrhage

GLCM : Gray level co-occurrence matrix

DCT : Discrete Cosine Transform

DWT : Discrete wavelet transform

REPTree : Reduced error pruning tree

ROC : Receiver operating curve

2-D : 2- Dimension

ROI : Region of Interest

SVM : Support vector machine.

MDEE : Multilevel Dominant Eigenvector Estimation

BDM : Bhattacharyya Distance Measure

DW : Diffusion weighted

KNN : K- Nearest neighbour

HFS : Hybrid feature selection

FV : Feature vector

Declarations section

Availability of data and materials

The data that support the findings of this study are openly available in Kaggle Repository at

https://www.kaggle.com/afridirahman/intracranial-brain-hemorrhage-ct-images

Competing interests

The authors declare that they have no competing interests.

Funding

Not applicable.

Authors' contributions

Santwana Gudadhe and Anuradha Thakare came up with the idea of the work and implemented the proposed method.Experiments, manuscript drafting and writing done by both the authors.All authors read and approved the final manuscript.

Corresponding Author

Correspondence to Santwana Gudadhe

Acknowledgement

No additional acknowledgements.

Authors' information

Santwana Gudadhe is currently pursuing the Ph.D. degree in Computer Engineering, from Savitribai Phule Pune University. She has received the B.E. degree in Information Technology and M.E. degree in Computer Science. Her research interests include image processing, deep learning, computer vision, Medical Imaging, Machine Learning and AI.

Affiliation

Department of Computer Engineering, Pimpri Chinchwad College of Engineering, Pune, INDIA.

Anuradha Thakare completed her PhD in Computer Science and Engineering from SGB Amravati University. She is currently working as a Professor in the Computer Engineering Department, PCCoE, Pune. She has received the B.E. degree in and M.E. degree in Computer Science. Her research interests include Machine Learning, Evolutionary Computing, Artificial Intelligence, Data Analytics and Medical Imaging.

Affiliation

Department of Computer Engineering, Pimpri Chinchwad College of Engineering, Pune, INDIA.

References

1. Marcolini, E.; Stretz, C.; DeWitt, K.M. Intracranial Hemorrhage and Intracranial Hypertension. Emerg. Med.Clin. N. Am. 2019, 37, 529–544.
2. Castro, J.; Chabert, S.; Saavedra, C.; Salas, R. Convolutional neural networks for detection of intracranial hemorrhage in CT images. In Proceedings of the 5th Congress on Robotics and

Neuroscience, Valparaíso, Chile, 27–29 February 2020; Volume 2564.

3. Agata Sage, Paweł Badura. Intracranial Hemorrhage Detection in Head CT Using Double-Branch Convolutional Neural Network, Support Vector Machine, and Random Forest. Appl. Sci. 2020, 10, 7577;
4. Nawabi J, Kniep H, Kabiri R, Broocks G, Faizy TD, Thaler C, Schön G, Fiehler J and Hanning U (2020). Neoplastic and Non-neoplastic Acute Intracerebral Hemorrhage in CT Brain Scans: Machine Learning-Based Prediction Using Radiomic Image Features. Front. Neurol. 11:285. DOI: 10.3389/fneur.2020.00285.
5. A. Aafreen Nawresh, S. Sasikala. An Approach for efficient classification of CT scan Brain Hemorrhage types using GLCM features with multilayer perceptron. ICDSMLA 2019 (pp.400-412), Springer Nature.
6. A.Padma, R. Sukanesh. SVM Based Classification of Soft Tissues in Brain CT Images Using Wavelet-Based Dominant Gray Level Run Length Texture Features. Middle-East Journal of Scientific Research 13 (7): 883-888, 2013 ISSN 1990-9233
7. Mahmoud Al-Ayyoub, Duaa Alawad, Khaldun Al-Darabsah, Inad Aljarrah. Automatic Detection and Classification of Brain Hemorrhages. WSEAS TRANSACTIONS on COMPUTERS, E-ISSN: 2224-2872, Issue 10, Volume 12, October 2013.
8. Shanbhag, S., Udupi, G. and Ranganath, K. 2018. Computer Based Method to Automatically Identify the Regions of Intracerebral Haemorrhage on DiffusionWeighted Magnetic Resonance Images. European Journal of Engineering and Technology Research. 3, 11 (Nov. 2018), 89-94.
9. D. Bhargava, S. Vyas, Ayushi Bansal. Comparative analysis of classification techniques for brain magnetic resonance imaging images, Advances in Computational Techniques for Biomedical Image Analysis. Academic Press,2020, Pages 133-144, ISBN 9780128200247.
10. A. Aafreen Nawresh, S. Sasikala. An Approach for Efficient Classification of CT Scan Brain Haemorrhage Types Using GLCM Features with Multilayer Perceptron, ICDSMLA 2019, Lecture Notes in Electrical Engineering 601, Springer Nature 2020

11. Supriya S. Shanbhag, G. R. Udupi, and K. Ranganath, Computer Based Method to Automatically Identify the Regions of Intracerebral Haemorrhage on Diffusion-Weighted Magnetic Resonance Images, EJERS, European Journal of Engineering Research and Science Vol. 3, No. 11, November 2018.
12. K Manikantan , Vaishnavi Govindarajan , V V S Sasi Kiran , S Ramachandran. Face Recognition using Block-Based DCT Feature Extraction, Journal of Advanced Computer Science and Technology, 1 (4) (2012) 266-283.
13. Shaveta Thakral , Pratima Manhas. Image Processing by Using Different Types of Discrete Wavelet Transform, Second International Conference, Advanced Informatics for Computing Research 2018.
14. (2008) Discrete Wavelet Transform (DWT). In: Furht B. (eds) Encyclopedia of Multimedia. Springer, Boston, MA. https://doi.org/10.1007/978-0-387-78414-4_305
15. https://www.kaggle.com/afridirahman/intracranial-brain-hemorrhage-ct-images (Last Referred on 25th October 2021).
16. Vrushali Y Kulkarni, Pradeep K Sinha, Random Forest Classifiers: A Survey and Future Research Directions, International Journal of Advanced Computing, ISSN:2051-0845, Vol.36, Issue.1, 2013.
17. Ajay Kumar Mishra, Bikram Kesari Ratha. Study of Random Tree and Random Forest Data Mining Algorithms for Microarray Data Analysis, International Journal on Advanced Electrical and Computer Engineering (IJAECE), X Volume -3, Issue -4, 2016.
18. Yas A. Alsultanny, Machine Learning by Data Mining REPTree and M5P for Predicating Novel Information for PM10, Cloud Computing and Data Science, Volume 1 Issue 1 2020.