CRITICAL ANALYSIS OF MACHINE LEARNING VS DEEP LEARNING ALGORITHMS FOR SMALL AND LARGE DATASETS.

Adithya Varambally, Akshaya S, Mukesh Madavi, Mohammed Mughees, Dheeraj D

Information Science and Engineering.
Global academy of technology, Bangalore, India.

Abstract : Image Classification is an important task within the field of computer vision. Image classification refers to the labelling of images into one of a number of predefined categories. Classification includes image sensors, image pre-processing, object detection, object segmentation, feature extraction and object classification. Image classification is an important and challenging task in various application domains, including biomedical imaging, biometry, video surveillance, vehicle navigation, industrial visual inspection, robot navigation, and remote sensing. Many Machine Learning and Deep Learning classification techniques like Artificial Neural Network(ANN), Decision Tree(DT), Support Vector Machine(SVM) ,Convolutional Neural Network(CNN), ResNet etc. have been developed for image classification. In this survey various Machine Learning and Deep Learning classification techniques are applied on various types of datasets. Machine learning algorithms generally give efficient and accurate result when applied on lighter datasets where as for heavier datasets Deep Learning algorithms are used.

In our survey we will analyze the algorithms like ResNet, convolutional neural network(CNN), Bagging, Bayesian, Decision trees(ID3), Support vector machines, Discriminant Analysis, Nearest Neighbors, Neural network, Random forests by applying them on various datasets to check if the algorithms provide the accurate result as expected. By doing the comparison of the accuracy obtained by applying all the algorithms on a particular dataset we get to know the most efficient algorithm that can be applied on that particular dataset for image classification. We also try to increase the size of the dataset and apply the algorithm to check if the algorithm performs as predicted, this is considered as critical analysis of the algorithm. Suppose an algorithm does not provide the predicted accuracy, we tweak the parameters that led to the decrease in the accuracy, this is considered as optimization of algorithms.

I. INTRODUCTION

In the area of computer vision image classification is an important task. Image classification usually labels the images into predefined categories. Classification includes image pre-processing followed by object detection, object segmentation and mainly feature extraction and object classification. Image classification has a lot of significance and has challenging task in various domains including biomedical imaging, video surveillance, vehicle navigation, remote sensing, biometry and navigation of robot. Many Machine Learning and Deep Learning classification techniques like Artificial Neural Network(ANN), Decision Tree(DT), Support Vector Machine(SVM) ,Convolutional Neural Network(CNN), ResNet etc.

Many supervised and unsupervised Deep learning and Machine learning algorithms are used for classification of images that belong to application filed and medical filed. Usually lighter datasets are used for image classification in application field where as comparatively larger and heavier datasets are used in case of medical field. Machine learning algorithms generally give efficient and accurate result when applied on lighter datasets where as for heavier datasets Deep Learning algorithms are used.

CLASSIFICATION USING DEEP LEARNING NEURAL NETWORKS FOR BRAIN TUMORS (Heba Mohsen et.al proposed)

In this paper the author has used Deep Neural Network classifier which is one of the DL architectures for classifying a dataset of 66 brain MRIs into four different classes .Brain tumor usually occurs whenever abnormal or uncontrolled division of brain cells within the brain is observed. Brain Tumors are of two types namely malignant and benign. Benign tumors do not spread whereas malignant tumors are cancerous as they grow with undefined boundaries very fast. MRI is a brain imaging technique that is generally employed for detection of brain tumor as MRI scanning provides lot of information about the human brain. MR images are loaded to the computer to perform image classification. The author has applied the deep learning concept to perform brain tumors classification using brain magnetic resonance images and measure the performance. The major agenda of the method proposed is to differentiate between normal brain and brain affected by tumors of different types. The following method uses a set of features extracted aided by the discrete wavelet transform (DWT), to train the DNN classifier for classification of brain tumors. Out of various Deep Learning architectures, the author has preferred to use convolutional neural networks (CNN) that can perform numerous sorts of complex operations using convolution filters. After the extraction of the features, the classification step using deep neural network is performed on the obtained feature vector. Classification is done by DNN containing 7 hidden layers structure. The author claims to have obtained the accuracy of 96.97%. The author claims that the good results that were obtained using the DWT could be employed with the Convolutional Neural Network in the future.

A METHOD FOR MEDULLOBLASTOMA TUMOR DIFFERENTIATION BASED ON CONVOLUTIONAL NEURAL NETWORKS AND TRANSFER LEARNING(Angel Cruz-Roaa et.al proposed)

In this paper, the author has used the approach of transfer learning for image classification. In transfer learning, no dataset containing the images to be classified is used as such. Instead we make use of the pre-trained models which are used either as conventional features or as a weight initialization strategy for a neural network which is retrained for a different task. AlexNet, Visual Geometry Group (VGG) and Over feat are the most popular pre-trained CNN models that are publicly available and have won the ImageNet challenge, held every year from 2012 onward. The ImageNet dataset contains millions of annotated images that has its place in over one thousand different categories. Medulloblastoma is a type of malignant brain tumors, it comprises around 25 percent of all brain tumors observed in children. Depending on the subtype of medulloblastoma i.e, anaplastic or non-anaplastic,the predicted course of tumor tends to differ, the anaplastic tumor typically last longer. A decision support tool for pathologists to help differentiate the anaplastic from the non-anaplastic tumor would aid in improved and efficient prognosis prediction and aid in de-escalation of therapy for the less aggressive tumors.The approach followed by the author is to compare two different CNN models VGG-CNN and IBCa-CNN, that were previously trained in two different domains namely, natural and histopathology images. Both CNNs are used as feature extractors to represent the content of the histopathology image regions for both the anaplastic and non-anaplastic medulloblastoma tumors from whole-slide images.

VGG-CNN is a CNN model trained to classify natural images in 1,000 categories. This CNN has around 138 million of parameters distributed in16 layers where 13 layers are convolutional-pooling layers, 2 are fully-connected layers and 1 layer is softmax classification layer. IBCa-CNN is a CNN model trained to classify histopathology images between invasiveor non-invasive breast cancer. IBCa-CNN has 3 layers : one convolutional-pooling layer, one fully-connected layer and one softmax classification layer.

The final step consists of training a classifier model using VGG-CNN or IBCa-CNN to classify anew histopathology image region that was unseen, as anaplastic or non-anaplastic. Result showed that IBCa-CNN outperformed VGG-CNN, despite being a smaller architecture. IBCa-CNN was able to learn features which are characteristic of histopathology images. Hence, histopathology signatures learnt from breast cancer pathology was also useful for medulloblastoma differentiation.

The author claims that VGG-CNN trained model could not classify histopathology images of medulloblastoma images whereas in IBCa-CNN trained model, the average classification accuracy was 89.8%, with a standard deviation of 5.6%.

APPLICATION OF DEEP TRANSFER LEARNING FOR AUTOMATED BRAIN ABNORMALITY CLASSIFICATION USING MR IMAGES(MuhammedTalo et.al proposed)

Magnetic resonance imaging (MRI) is the most common technique used to detect brain tumors. Traditionally, MRI images are manually analyzed by radiologists in order to detect the abnormal brain conditions. Manual analysis of huge volume of images is time consuming and tedious. Hence, detection using computer helps in accurate and fast diagnosis. In this paper, the author uses an approach of deep transfer learning to automatically classify brain tumors using MR images. Here deep learning model used is a CNN based ResNet34 model. The author has used deep learning techniques such as data augmentation, optimal learning ratefinder and fine-tuning in order to train the model. The proposed model achieved 5-fold classification and the author claims that the accuracy obtained was of 100% when the dataset consisting of 613 MR images was used for classification. Abnormalities like Alzheimer's disease, stroke, Parkinson's disease and autism were classified by ResNet34.

Here previously trained model that have been learned how to solve a similar classification problem. The ResNet34 architecture is trained on ImageNet database which contains more than one million images that belong to 1000 categories.

The author proposes that ResNet34 architecture converges faster when compared to other pre-trained models such as VGG and inception. The ResNet34 architectureis very simple to use in various datasets when compared to other pre-trained models suchas inception andVGG.

Layer (type)	Input Shape	Output Shape	Number of Parameters
ResNet34 (Model)	(3, 128, 128)	(64, 64, 64)	9408

FULLY AUTOMATIC SEGMENTATION OF BRAIN TUMOR IMAGES USING SUPPORT VECTOR MACHINE CLASSIFICATION IN COMBINATION WITH HIERARCHICAL CONDITION RANDOM FIELD REGULARIZATION. (Stefan Bauer et.al proposed)

The author has used Support Vector Machine classification algorithm applying multispectral intensities and the pattern with the subsequent stratified regularization based on the Conditional Random fields to propose a fully automatic system for brain tissue classification and segmentation. The essential task to analyze a brain cancer is to delineate the tumor boundaries from MRI. Hence, the first task is to separate tumor tissues and healthy tissues before both the regions are sub classified into white matter, active, gray matter and necrotic, edema region and cerebrospinal fluid in a novel hierarchical way. This approach is fast and also adapted to standard clinical acquisition protocols. The author has considered the case of glioma, since it is the most aggressive type of brain tumors. The conditional random field approach makes the process very efficient but also maintains accuracy by considering the neighbor relationships.

The prominent features for distinguishing healthy tissues and pathological and also their sub regions were extracted from the multispectral imaging data. The first order texture features such as energy, variance, mean, kurtosis, entropy and skewness were also used. Dice similarity coefficient was used to evaluate the results. For intra patient case, small sub regions were trained on the classifier and for inter patient case, leave-one-out cross-validation was performed. They obtained the results as such:

	GTV	Necrotic	Active	Edema
Intrapatient Regularized	**0.84 ± 0.03**	0.61 ± 0.24	0.71 ± 0.09	0.73 ± 0.04
Intrapatient Unregularized	**0.76 ± 0.10**	0.45 ± 0.31	0.59 ± 0.16	0.72 ± 0.07
Interpatient Regularized	**0.77 ± 0.09**	0.45 ± 0.23	0.64 ± 0.13	0.60 ± 0.16
Interpatient Unregularized	**0.67 ± 0.13**	0.30 ± 0.24	0.46 ± 0.12	0.63 ± 0.16

MACHINE LEARNING AND DEEP LEARNING TECHNIQUES TO PREDICT OVERALL SURVIVAL OF BRAIN TUMOR PATIENTS USING MRI IMAGES (LinaChato et.al proposed)

In this paper the author has used MRI images of the patients' to predict the survival rate with glioma brain tumor. Glioma is considered as the aggressive type of brain tumor and overall survival rate doesn't exceed two years and they represent 74.6% of all malignant tumors. Multimodal brain tumor segmentation 2017 challenge dataset was used, which has 163 samples. The sample contained patient's age, MRI brain images, and the overall survival time in days. The dataset was labeled according to the survival rate i.e. short-term survivors (less than 6 months), mid-term survivors (ranging from 10-15 months), and long-term survivors (more than 15 months). The dataset contained two groups, higher grade glioma and lower grade glioma and four sequences of MRI modalities were provided.

Since the size of the dataset was small, various Machine Learning methods such as K-Nearest Neighbors, Support Machine Vector, Tree, Ensemble, Linear Discriminant, and Logistic Regression were used to develop the prediction model for classification. Many feature extraction methods were used such as Statistical and Intensity Texture, Volumetric and Location, 2D Deep Feature, and Histogram Distribution.

The first order features includes Correlation, Energy, Standard Deviation, Smoothness, Entropy, Root Mean Square, Contrast, Homogeneity, Inverse Difference Moment, and Mean. The best prediction accuracy based on classification is achieved by using deep learning features extracted by a pre-trained Convolutional Neural Network and was trained by a linear Discriminant. The author obtained an accuracy of 68.5%.

A COMPARATIVE STUDY OF DIFFERENT MACHINE LEARNING METHODS ON MICROARRAY GENE EXPRESSION DATA (MEHDI PIROOZNIA et. Al proposed)

In this paper, the author have compared the efficiency of many classification methods of Machine Learning algorithms and Deep Learning algorithms such as Artificial Neural Network, Support Vector Machine, Radial Basis Function Neural Networks, Multilayer Perceptron Neural Network, Decision Tree, Bayesian, and Random Forest on Micro Array Gene Expression dataset. To calculate the accuracy of these Machine Learning classifiers, v-fold cross validation was used and also common clustering methods were applied to analyze the efficiency such as Discriminately Boosted Clustering, k-means clustering and expectation-maximization clustering. The author had also compared the efficiency of many feature selection methods such as Chi Squared, Support Vector Machine Recursive Feature Elimination, and CFS. These methods were applied to eight different binary micro array-datasets..

Dataset	Comparison	Variables (Genes)	Samples
1. Lymphoma (Devos et.al, 2002)	Tumor vs. Normal	7129	25
2. Breast Cancer (Perou et. al, 2000)	Tumor subtype vs. Normal	1753	84
3. Colon Cancer (Alon et. al, 1999)	Epithelial vs. Tumor	7464	45
4. Lung Cancer (Garber et. al, 2001)	Tumor vs. Normal	917	72
5. Adenocarcinoma (Beer et.al, 2002)	NP vs. NN	5377	86
6. Lymphoma (Alizadeh et al, 2000)	DLBCL1 vs. DLBCL2	4027	96
7. Melanoma (Bittner et. al, 2000)	Tumor vs. Normal	8067	38
8. Ovarian Cancer (Welsh et. al, 2001)	Tumor vs. Normal	7129	39

The percentage of accuracy obtained after applying cross validation were as such:

Dataset	SVM	RBF Neural Nets	MLP Neural Nets	Bayesian	J48 Decision Tree	Random Forest	Id3	Bagging
1. Lymphoma (Devos et.al, 2002)	96.0	84.0	68.0	88.0	64.0	76.0	48.0	52.0
2. Breast Cancer (Perou et. al, 2000)	97.6	97.6	96.4	92.9	92.9	96.4	94.0	96.4
3. Colon Cancer (Alon et. al, 1999)	95.6	91.1	91.1	93.3	91.1	80.0	88.9	93.3
4. Lung Cancer (Garber et. al, 2001)	97.2	97.2	97.2	95.8	94.4	95.8	97.2	97.2
5. Adenocarcinoma (Beer et.al, 2002)	96.5	94.2	75.6	75.6	74.4	79.1	66.3	79.1
6. Lymphoma (Alizadeh et al, 2000)	96.9	88.5	75.0	85.4	75.0	76.0	62.5	84.4
7. Melanoma (Bittner et. al, 2000)	94.7	81.6	84.2	76.3	81.6	81.6	52.6	81.6
8. Ovarian Cancer (Welsh et. al, 2001)	94.9	84.6	89.7	87.2	87.2	89.7	74.4	89.7

A MACHINE LEARNING APPROACH TO BRAIN TUMORS SEGMENTATION USING ADAPTIVE RANDOM FOREST ALGORITHM.(OmidReyhani-Galangashi et.al proposed)

In this paper the authors, have used a brain tumor segmentation method based on the Random Forest algorithm. The method proposed is applied to the part of the images of brain magnetic resonance and a high performance indices like Dice Similarity Coefficient (DSC) as well as algorithm accuracy (ACC) are calculated which are 98.38% and 97.65%, respectively. The result which obtained, showed that the proposed model can have a good performance when compared to the other segmentation methods. In order to find the exact area of the brain tumors, a cohesion-based self-merging (CSM) algorithm for the division of the brain MRI data is proposed. CSM attracts a lot of attention because it has more favorable results than that of other integration processes. In the proposed work, the noise effect is greatly reduced to increase the chance of finding the exact area of the tumor ,so that the calculation time is also a lot lower ,since the algorithm proposed is simple and therefore computes with less complexity. So a particle swarm optimization (PSO) algorithm which is based on clustering is proposed and identifies the center of mass of clusters. Each cluster contains the brain tumor patterns obtained by MRI in a group. The results of the three different performance measurements were compared with the results obtained by Support Vector Machine (SVM) and Adaptive Boosting (AdaBoost) methods. Performance analysis showed that the qualitative results of the proposed model are similar to those obtained with other two models. In addition, various values of the particle swarm optimization (PSO) control parameters are selected to obtain better results from the algorithm. At first, the system proposed recognizes the MRI tumor by using the naïve bayes classification. The mechanism of the proposed method is that at first the brain image is divided, and then the suspected area is identified according to the symmetric plate and fuzzy classification for the tumors. While performing the division step, the tumor area was successfully identified using the variable model combination and spatial relationships. After doing diagnosis, clustering methods and boundaries were used to extract the exact tumor area and above 99% accuracy was obtained for diagnosis—according to the authors' claim. The author's approach was to manipulate machine learning and customize an innovative learning algorithm known as Random Forest (RF) in a way to handle larger datasets. In this regard, after presenting the mathematical modeling of the RF concept, the author suggested a work flow to implement the segmentation process on tumors for MRI data. In this research, the author applied a proposed method to extract features from the MRI data. The features include first-order, higher-order and texture that have been utilized for the purpose of reducing the size of the processed data.

EFFICIENT KNN CLASSIFICATION WITH DIFFERENT NUMBER OF NEAREST NEIGHBORS.(Shichao Zhang et.al proposed)

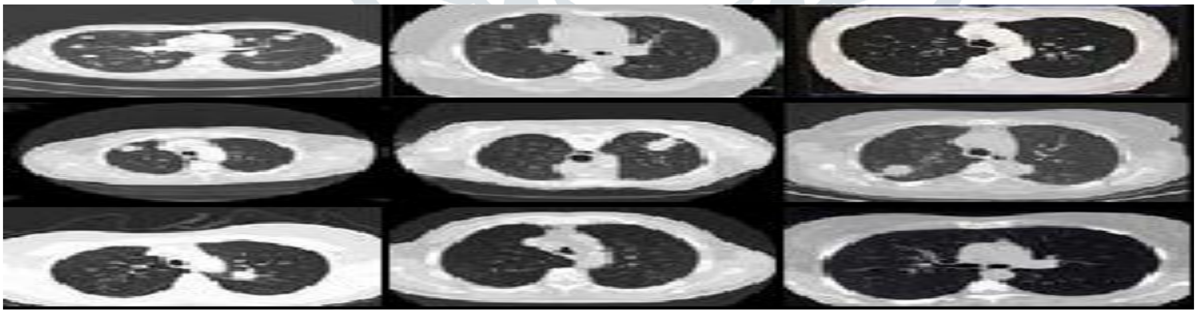
In this paper, the author has used a kTree method to learn different optimal k values for different test/new samples, by involving a training stage in the kNN classification. Particularly, in training stage, this strategy learns ideal k-values for training samples utilizing a sparse recreation model and afterward continues to make a decision tree (specifically, kTree) by utilizing preparing tests and the ideal k values learnt. In the test stage, the kTree outputs the optimal k value for each test sample, and then, the kNN classification is conducted using the learned optimal k value and all training samples.

As a result, the proposed method has a same running cost but more classification accuracy, compared tol kNN methods used traditionally, where all test samples a fixed k value is assigned to. Also, the proposed technique takes less running expense yet accomplishes a decent classification accuracy, relatively to the new kNN strategies,which to different test samples, assign a different k values. This paper further says an improved type of kTree method (k*Tree method) to fasten its test stage by storing extra information of the training samples in the kTree's leaf nodes , such as the training samples in the leaf nodes, their kNNs, and the neighbor close to these kNNs. The key idea of the proposed methods is to design a training stage for reducing the running cost of test stage and improving the classification performance. 20 public data sets from UCI Repository of Machine Learning Data have been conducted to evaluate the proposed methods with the competing methods, and the experimental results indicated that our methods outperformed the competing methods in terms of classification accuracy and running cost. In future, we will focus on improving the performance of the proposed methods on high-dimensional data.

AN ENHANCED K NEAREST NEIGHBOR METHOD TO DETECTING AND CLASSIFYING MRI LUNG CANCER IMAGES FOR LARGE AMOUNT DATA (P. Thamilselvan et.al proposed)

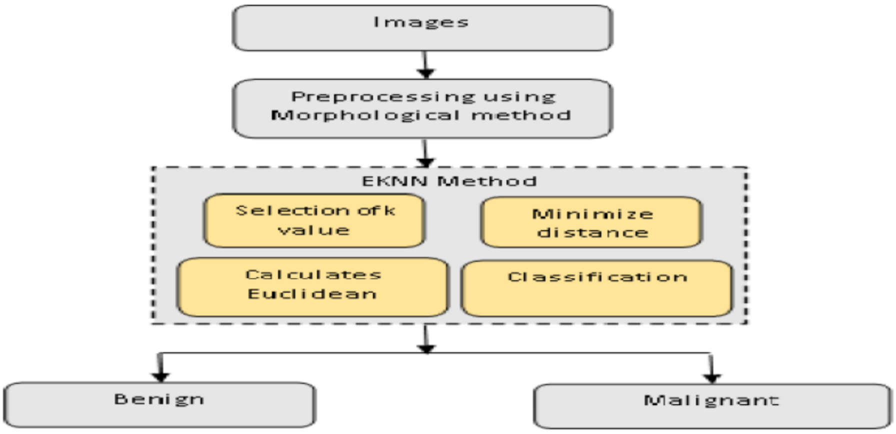
We have used enhanced K Nearest Neighbor (EKNN) classify the cancer in this method as to be a benign or malignant cancer tissues based on the large lung cancer image dataset tested. In this method we first start with data-preprocessing, later we identify the type of cancer the dataset is and then classify it. General morphological method has been used in preprocessing phase. For identifying and classifying the cancer we have used enhanced k nearest algorithm. For distance measure we have used Euclidean distance where we use 4 steps of EKNN to classify the cancer.

In this proposed work the MRI images of the part of body to be tested for cancer is taken. 256x256 dimension of region has been taken from the intended images. Here are the few lung cancer MRI image sample dataset.



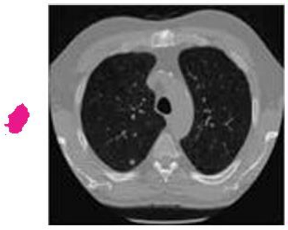
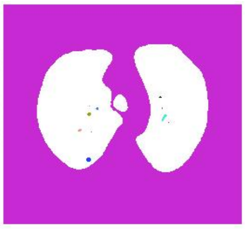
General morphological method has been used in preprocessing the images where noisy imagers are filtered followed by separation of benign cancer from malignant cancer tissues. Dilation and erosion is the general morphological method we have used in this work to preprocess the images. Removing noises is done by dilation process. Converting original image to black image where the size of width is reduced by erosion process.

The following diagram shows the system of the work



In enhanced K nearest algorithm we first choose k value (k value is user's preference). We use Euclidean distance for distance measure. We store all training dataset A=(**a**i,**b**i)**,** where **a**i is training dataset and **b**i is coherent class. We then step forward to testing phase where we calculate distance between all training data and new feature vector. Minimum k value is then derived after arranging the acquired distance one by one. After this algorithm is used in the preprocessed images classification and identification of cancer tissues are done.

Following images are cancer detected images

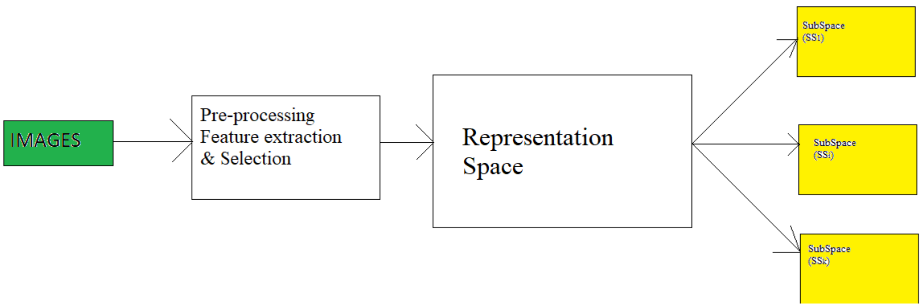
Result obtained by this method is shown below

Proposed method	Classification accuracy	Processing time	Minimum neighboring distance	Misclassification rates
EKNN	97%	3sec	0.20889	3%

IMAGE CLASSIFICATION USING SUPPORT VECTOR MACHINE AND ARTIFICIAL NEURAL NETWORK

Here we have used Artificial neural network(ANN) and support vector machine(SVM) to classify images. We first do separation of an image into many sub-images on the basis of its features. These sub-images are divided into responsive classes by ANN and then SVM is applied on these responsive classes.

After preprocessing and feature extraction are done images would present in large representation space. Dimension reduction is done and it will be processed into sub-spaces



After the subspaces are made feature vector would be extracted from that image. This feature vector is input to the ANN . The number of nodes in input layer will be the number of feature vector called in.

There will be k number of classification results of sub-space if there are k subspaces. It is integrated and mean value is taken.

$$\mathrm{CL} = \frac{1}{k}\sum_{i=1}^{k} \mathrm{w_L\ CL\_SS_L}$$

Where wi is the classification's weight, which is result of subspace SSi, and satisfies:

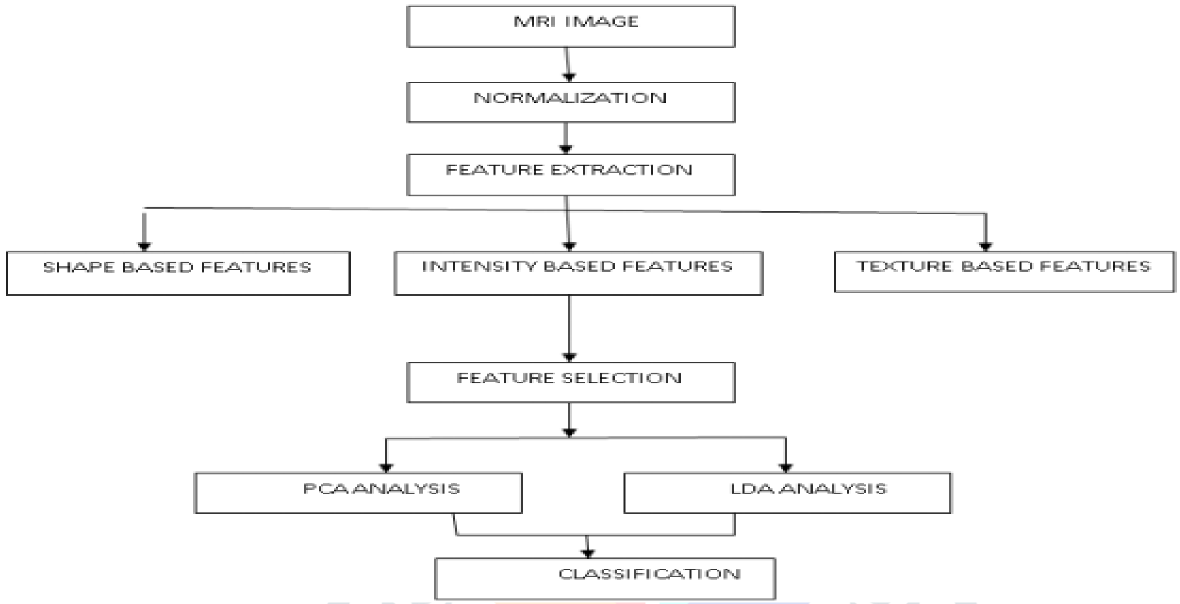
$$\sum_{i=1}^{R} w_i = 1$$

We combined all ANNs results by using SVM and thus SVM is the solution for identifying the weights of ANNs results.

By using this method the average classification rate is 86%.

BRAIN TUMOR MRI IMAGE CLASSIFICATION WITH FEATURE SELECTION AND EXTRACTION USING LINEAR DISCRIMINANT ANALYSIS(V.P. GladisPushpa Rathiet.al proposed)

This method is an innovative way of feature extraction and selection. This method focuses on using many shapes, texture and many more feature of tumor as white matter, gray matter, CSF, abnormal and normal area. We have used Principal Component Analysis (PCA) and Linear Discriminant Analysis (LDA) and also Support Vector Machines (SVM) in this method. Comparison of linear techniques and non-linear techniques are done by SVM. PCA and LDA are used in dimension reduction. The architecture of our proposed work is as follows



Images are taken from different patients with gliomas. 24 slices in axial plain with 5 mm slice thickness in each volume. MR imaging was done in 3T siemens devices. MRI image dataset is done as shown in the following (normalization done in range of 0-255)

Attribute	Description	Value
Age	Age in Years	17 to 83
Sex	Sex	Men -46, Women -52
Matrix size	Size of the matrix	192*256*192
Voxel size	Size of the voxel	0.98*0.98*1mm
Sequences	MRI image sequences	Axial 3D T1 weighted , Sagittal 3D T2 weighted , Fluid Attenuated Inversion Recovery (FLAIR)

Normalization is done by converting the images to gray levels of 0-1 and features extraction is done.
Feature Extraction here is done based on shape, intensity and texture.

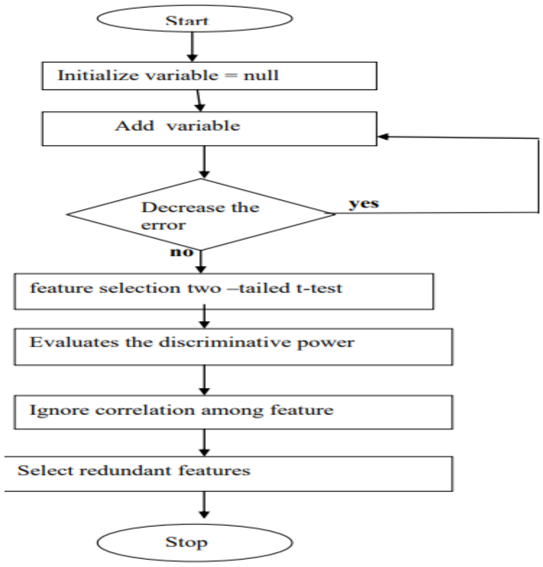
Shape Features – shape being circular, regularity, Area.

Intensity features – Mean of the data, distortion , standard deviation etc..

features –Contrast, information measure, unifromity, cluster attributes like shades, sum of standard deviation

Feature selection is done where the extracted images with least dimensions are taken and rest are discarded. It is done for the data to run on the algorithms smoothly.

Forward Selection starts with no variables and iteratively ads one by one which makes least errors. We use simple rank-based selection criterion, which compares two distributions



Redundant values can be selected but not preferably. Feature ranking method can be used to choose more discriminative feature.

Backward Selection starts with all variables and deletes them iteratively until any other removal increases error significantly. This is done to reduce over fitting. To optimize the classification performance support vector machine recursive feature elimination algorithm is used. Backward selection makes SVM based error bound smallest.

Classification is done by PCA and LDA. Frequencies within class which are unequal are tested on randomly selected test data. Variance between classes and within are maximised in this method. The classification phase is divided into training and testing phase. Efficiency of training determines the accuracy of the classification.

This method's accuracy as compared to few others are as shown below

Features	T1	T2	FLAIR	TOTAL
Intensity	6	5	11	22
Shape	1	1	3	5
Texture	8	5	20	33
Total	10	20	30	60

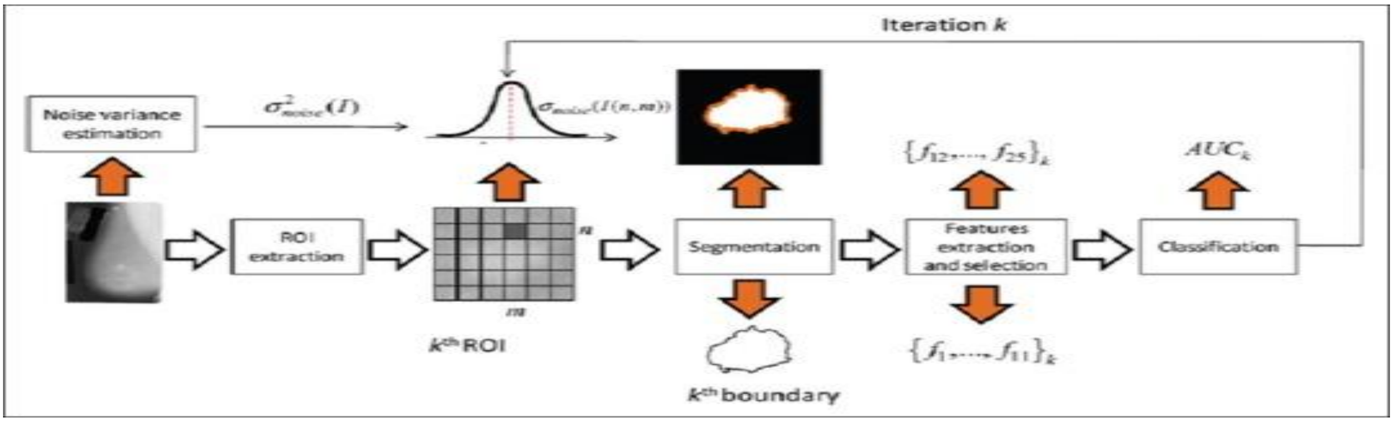
AN INVESTIGATION OF BAYES ALGORITHM AND NEURAL NETWORKS FOR IDENTIFYING THE BREAST CANCER(E UDAYAKUMAR ET.AL PROPOSED)

X-ray mammography is most used in method in early detection and screening of breast cancer, but the problem is in current mammogram, it doesn't provide consistent results for the radiologists. In this low-level processing techniques and image segmentation digital images of mammography can be lightened or darkened before they are printed; and thus, breast cancer detection can be done.

We have used naïve Bayes algorithm and neural network in this work to find the type of mammogram and stages. Gray-level co-occurrence matrix (GLCM) and texture feature are used for feature extraction in this work. To find the affected portion(region of interest) we do segmentation process using region-growing algorithm.

Block diagram of proposed system

After mass boundary which is the feature to be extracted is seen the structure is computed to simplify the large set of data accurately. Each detected mass can be indexed based on its danger level based on its classified feature. The steps of validation is done by computing the characteristic curve. Images are classified uniquely and then image segmentation is done.



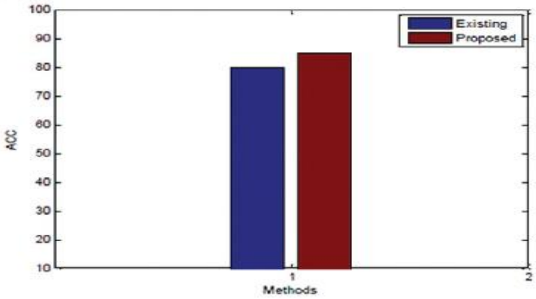
pepper and salt noise is removed and Gray-level co-occurrence matrix is used in preprocessing. The input image is like the following

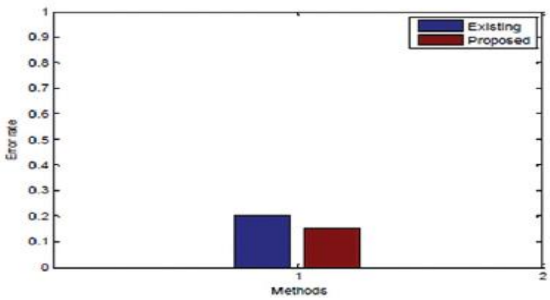
Histogram equalization method enhances the images that are taken in. Based on image luminosity level in the pixels of the image histogram levels tends to increase the contrast.

This enhanced image is shown here in the following

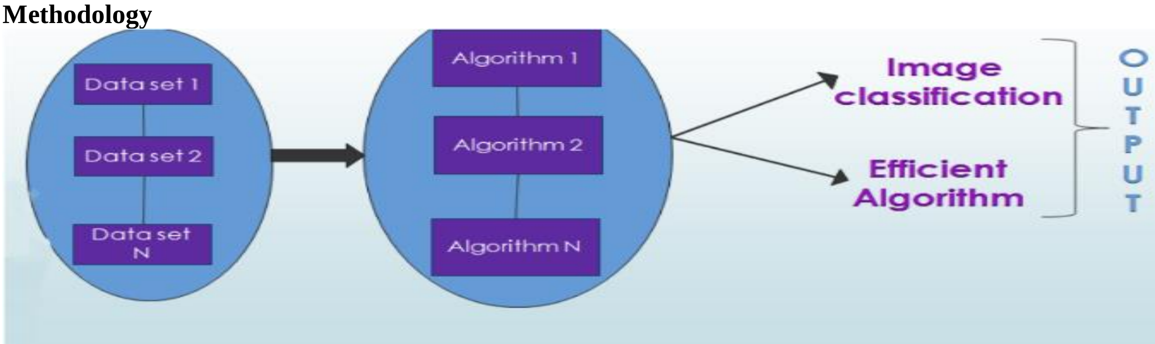
Image segmentation is done where partitioning of an image into multiple segments based on correlation between the neigbouringfeautres. The goal of segmentation is to simplify an image into something that is more meaningful and easier to analyse. Region-growing algorithm is used in locating the affected area here. It first takes a pixel here called as seed node and compares it to its neighbouring pixel and grows accordingly

After this images are classified visually based on the feature in segmented part of the image.Classification of unknown samples are done by Bayes algorithm and neural network. Accuracy of the proposed work as compared to existing mammogram is as shown below with error rate.





Methodology



We take different image classification algorithms and run different types of datasets on these different algorithms. We then check for its efficiency for all trials and record it.

Conclusion

We will take different image classification and use it on different datasets. We will check its efficiency and record it. We check for most optimised algorithms for different hyperparameters for different types of datasets to use.

References

[1]. Heba Mohsen, El-Sayed A. El-Dahshan, El-Sayed M. El-Horbaty, Abdel-Badeeh M. Salem , “Classification using deep learning neural networks for brain tumors “, IEEE, 2017

[2]. Angel Cruz-Roa, John Arevalo, Alexander Judkins, Anant Madabhushi, Fabio Gonzalez “A method for Medulloblastoma Tumor Differentiation based on Convolutional Neural Networks and Transfer Learning”, IEEE, 2016

[3]. Muhammed Talo, Ulas Baran Baloglu, Ozal Yildirim, U Rajendra Acharya “Application Of Deep Transfer Learning For Automated Brain Abnormality Classification Using MR Images”, IEEE, 2017

[4]. Stefan Bauer , Lutz-P.Nolte , Mauricio Reyes , “ Fully Automatic Segmentation of Brain Tumor Images Using Support Vector Machine Classification in Combination with Heierarchical Condition Random Field Regularization” ,IEEE 2011

[5]. Lina Chato , Shahram Latifi , Lina Saeed Chato , “ Machine Learning and Deep Learning Techniques to Predict Overall Survival of Brain Tumor Patients using MRI Images” , IEEE 2017

[6]. Mehdi pirooznia , Jack Y Yang ,Mary Qu Yang and Youping Deng , “A comparative study of different machine learning methods on microarray gene expression data” , IEEE 2018

[7]. Toktam Hatami , Mohammad Hamghalam , Omid Reyhani-Galangashi, Sattar Mirzkuchaki , “ A Machine Learning Approach to Brain Tumors Segmentation Using Adaptive Random Forest Algorithm ”, IEEE 2019

[8]. Le Hoang Thai ,Tran Son Hai, Nguyen Thanh Thuy,” Image Classification using Support Vector Machine and Artificial Neural Network”, IEEE 2012

[9]. Shichao Zhanf , Xuelong Li , Ming Zong , Xiaofeng Zhu , Ruili Wang,”Efficient kNN Classification With Different Numbers of Nearest Neighbors”, IEEE 2017

[10]. P. Thamilselvan, Dr. J. G. R. Sathiaseelan, “An enhanced k nearest neighbor method to detecting and classifying MRI lung cancer images for large amount data” , IEEE 2016

[11]. E Udayakumar, S Santhi, P Vetrivelan, “An Investigation of Bayes Algorithm and Neural Networks for Identifying the Breast Cancer”, IEEE 2017

[12]. V.P.Gladis Pushpa Rathi, Dr.S.Palani “BRAIN TUMOR MRI IMAGE CLASSIFICATION WITH FEATURE SELECTION AND EXTRACTION USING LINEAR DISCRIMINANT ANALYSIS ”, IEEE 2015