FEATURE SELECTION AND CLASSIFICATION OF HUMAN BRAIN MRI

Contents

Motivation

- Brain Abnormality $\Rightarrow$ life threatening.
- Human experts
  - Time consuming
  - Subjective
  - Expensive .

Problem Statement

- To propose an optimum subset of features that can distinguish normal and abnormal Human Brain MRI scans.

- To introduce a system for classification based on the selected subset of features which is better than the already existing systems.

Contributions

(i) Different texture families are combined for MRI classification.

(ii) GA swarm intelligence ⇒ for feature selection, better than traditional statistical approaches.

(iii) Comparison between proposed classifier and existing research.

(iv) Research paper is about to be submitted.

Background & Literature Review

Features for Image Classification

Features

A set of variables believed to carry discriminating and characterizing information about the objects under consideration.

Common Types of Features

- Color
- Edge
- Shape
- **Texture**

  Most distinguishing feature *[1]

[1] J.Juntu, A. M. De Schepper,P.Van Dyck,D.VanDyck, J. Gielen, P.M. Parizel, and J.Sijbers, “Classification of Soft Tissue Tumors By Machine Learning Algorithms”. INTECH 2011

Why Texture Features?

- Texture information can improve accuracy of Brain MRI classification.[1-2]

- Combination of texture features from different families can lead to better classification performance .[3]

[1] J.Juntu, A. M. De Schepper,P.Van Dyck,D.VanDyck, J. Gielen, P.M. Parizel, and J.Sijbers, "Classification of Soft Tissue Tumors By Machine Learning Algorithms". 2011

[2] De Schepper, A., Vanhoenacker, F., Parizel, P. & Gielen, J. (eds) (2005). Imaging of Soft Tissue Tumors,3rd edn, Springer.

[3] M. A. García, D. Puig, "Improving Texture Pattern Recognition by Integration of Multiple Texture Feature Extraction Methods", pp. 7-10, 2002.

Texture Feature Families

- Du Buf et al [6] compared
  - GLCM
  - Fractal
  - Michell’
  - Knutsson’s
  - Laws’
  - Unser’s
  - curvilinear integration

  GLCM found to be best

[6] J.M.H. du Buf, M. Kardan, and M. Spann. Texture feature performance for image segmentation. Pattern Recognition, 23(3/4):291-309

Texture Feature Families

- Weszka et al [7] compared
  - Fourier power spectrum
  - GLCM
  - first-order statistics.

  GLCM found to be best
- R. Porter et al [8] compared
  - Wavelet Transform
  - Gabor Filter
  - Gaussian Markov Random Fields

  Wavelet found best

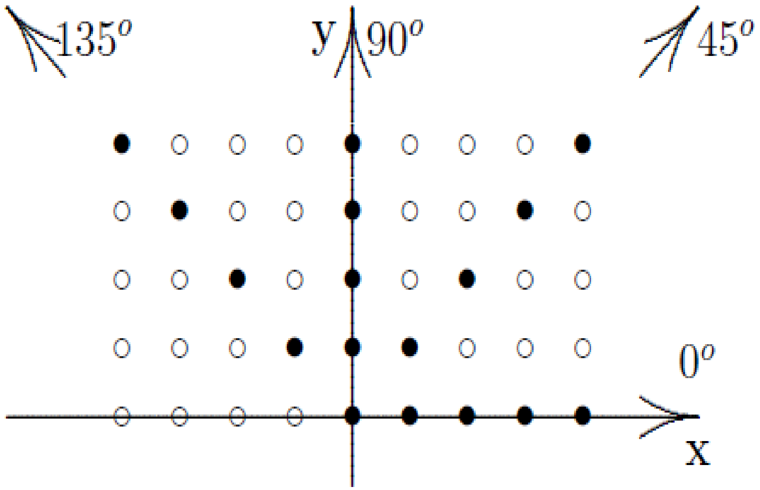
[7] J. Weszka, C. Dyer and A. Rosenfeld, A comparative study of texture measures for terrain classification, *IEEE Trans. Syst. Man. Cybernet. SMC-6, 269-285'(1976).*

[8] R. Porter and N. Canagarajah. Robust rotation-invariant texture classication: wavelet, gabor filter and gmrf based schemes. Vision, Image and Signal Processing, IEE Proceedings -, 144(3):180 {188, jun 1997.

Gray Level Co-occurrence Matrix (GLCM)

Contains a count of number of times a given feature occurs in a particular spatial relation to another feature.

- For a given distance *d* four angular GLCMs (θ = 0, 45, 90 and 135 degrees).



EXAMPLE

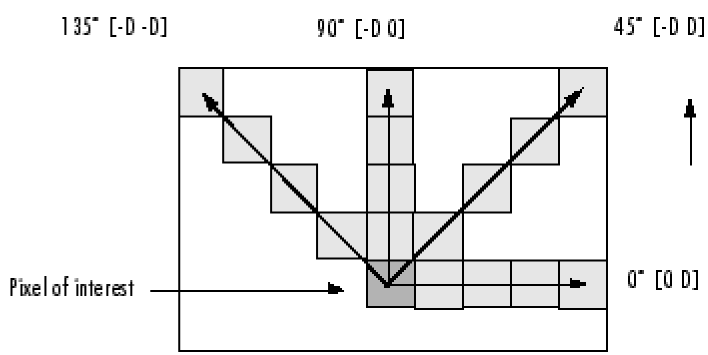
- Grey level co-occurrence matrix

I

1	1	5	6	8
2	3	5	7	1
4	5	7	1	2
8	5	1	2	5

GLCM

	1	2	3	4	5	6	7	8
1	1	2	0	0	1	0	0	0
2	0	0	1	0	1	0	0	0
3	0	0	0	0	1	0	0	0
4	0	0	0	0	1	0	0	0
5	1	0	0	0	0	1	2	0
6	0	0	0	0	0	0	0	1
7	2	0	0	0	0	0	0	0
8	0	0	0	0	1	0	0	0



GLCM Features

- GLCM Features, using d=1 and four angles (θ = 0, 45, 90 and 135 degrees).
- Calculating mean, range and variance of GLCM to avoid direction dependency (Haralick et al)
- 17 x 3 = 51 features

$$M_T(d) = \frac{1}{N_\theta} \sum_\theta T(d,\theta)$$

$$R_T(d) = \max_\theta[T(d,\theta)] - \min_\theta[T(d,\theta)]$$

$$V_T^2(d) = \frac{1}{N_\theta} \sum_\theta [T(d,\theta) - M_T(d)]^2$$

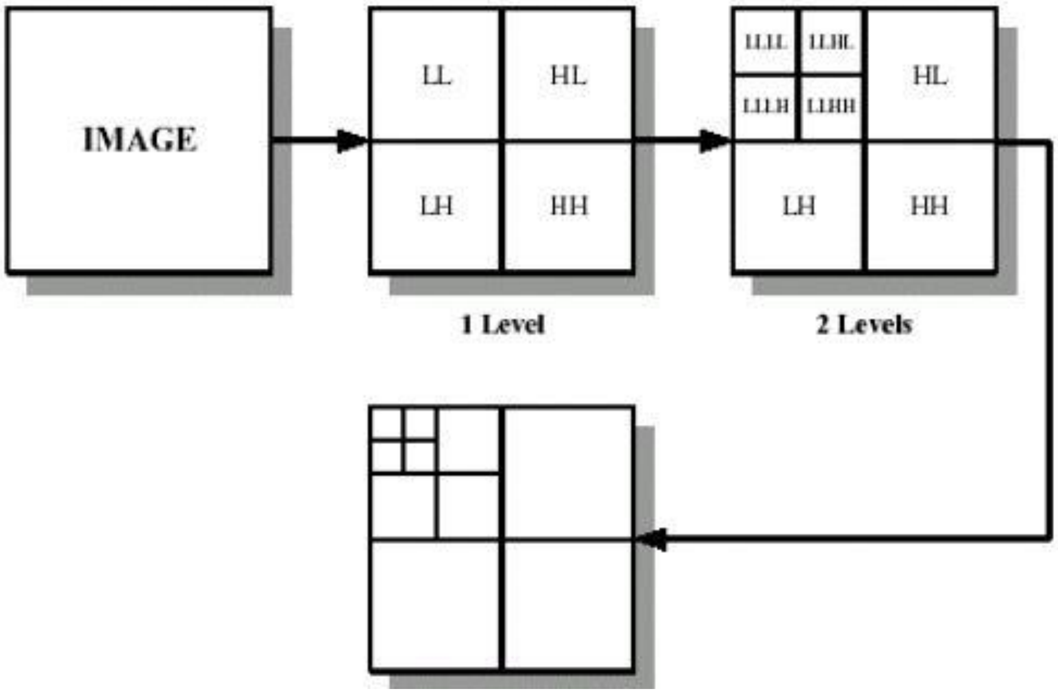
Wavelet Transform

- Provides a multi-scale analysis of an image.

- Information:
  - Horizontal(HL) details.
  - Vertical(LH) details.
  - Diagonal(HH) details.
  - Approximation(LL)

Wavelet Transform

- 4-channels or sub-bands per each scale of decomposition.
- Total of 12 features are extracted.



Feature Selection

Feature subset selection is a process of selecting a subset of features from a large number of features such that the selected features are powerful enough to discriminate effectively among different classes.

Methods for MRI feature selection

Meta-heuristic

A. Genetic Algorithm[5]
B. Particle Swarm Optimization.

[5] Garcia-Nieto J, Jourdan L., " A comparison of PSO and GA approaches for gene selection and classification of microarray data", Proc. Of the 9th Annual conference on Genetic and Evolutionary Computation 2007.

Genetic Algorithm

- Same behavior as natural selection of individuals.
- Based on the principle of Darwin theory "Survival of fittest".
- Used GA for the selection of optimal features.

Genetic algorithm

There are five phases

- Initial population
- Tournament selection
- Selection
- Crossover
- Mutation
- Fitness function

Initial population

- Begins with randomly generated states .
- These stats are satisfactory to the problem

```
1 1 0 0 0 ................ 0 0 1
1 0 1 1 0 .................1 0 1
1 1 0 1 0 .................0 0 1
1 0 0 1 0 .................1 1 0
.                             .
.                             .
.                             .
.                             .
.                             .

1 1 1 1 0 .................1 1 1
```

N*M
N=size of population
M=number of features

Fitness function

- The fitness function evaluates the fitness of individuals.
- A good fitness function should return better state.
- The fitness function give a score to each state.
- The probability of being chosen for reproduction depends on the fitness state.
- Fitness function KNN/SVM

selection

- Two pairs are selected at random to reproduce
- They are selected based on their fitness function score.
- One may be selected more than once or one may not be selected at all.

1 1 0 0 0 ....0 0 1: 80
1 0 1 1 0 ....1 0 1
. 70
. ....
1 1 0 1 0 ....0 0 1 60
1 1 1 1 0 ....1 1 1 90

Selected features
1 1 0 0 0 .... 0 0

1 1 1 1 0 ....1 1 1

Crossover

- Crossover causes different characteristics to be passed to the offspring from the parents.

p1: 0001001110010010

p2: 1010001001000011

After crossover

00010011 01000011 Note :crossover probability taken =0.5

10100010 10010010

Mutation

- Mutation = Growth of genes
- In our case, It means flipping of bits.

1000100110 Mutation prob = 0.2

0100100110

Flow chart:
GA for Feature Selection

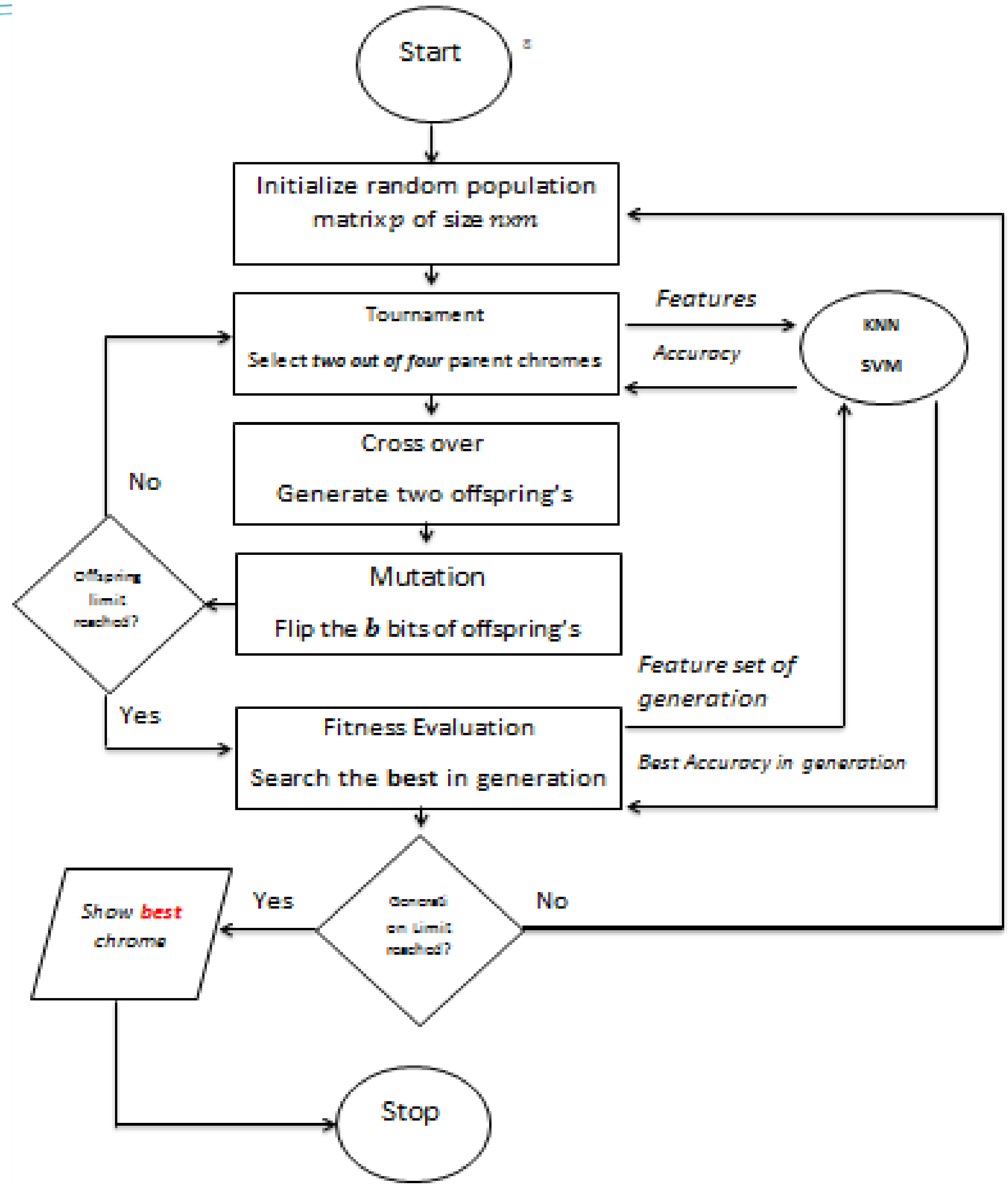


Figure 2: flow Chart of Feature selection (GA)

Machine Learning Classifiers

- Classifier: An algorithm which adjusts its parameters to find the correct decision boundaries –through a learning algorithm using a training dataset.
- Error: Incorrect labeling of the data by the classifier.
- Training Performance: The ability/performance of the classifier in correctly identifying the classes of the training data, which it has already seen.
- Generalization (Test Performance): The performance of the classifier in identifying the classes of previously unseen data.

Classifiers

- J.Juntu at el [1] compared
  - Neural Network Classifier
  - Decision trees Classifier
  - Parzen Classifier
  - KNN
  - SVM

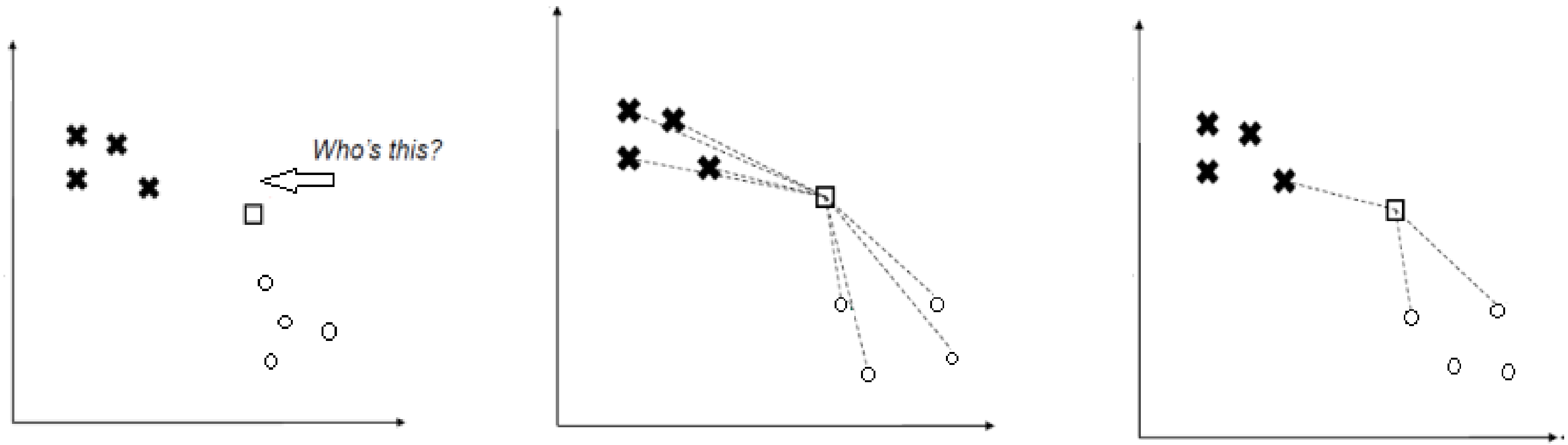
  SVM proved to be best

[1] J.Juntu, A. M. De Schepper,P.Van Dyck,D.VanDyck, J. Gielen, P.M. Parizel, and J.Sijbers, "Classification of Soft Tissue Tumors By Machine Learning Algorithms". 2011

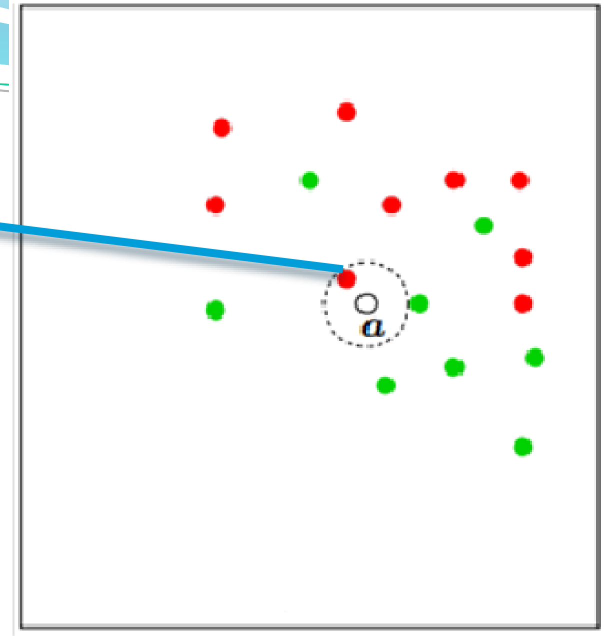
KNN Classifier

1. *Measure distance to all points.*
2. *Find closest "k" points.*
3. *Assign majority class.*

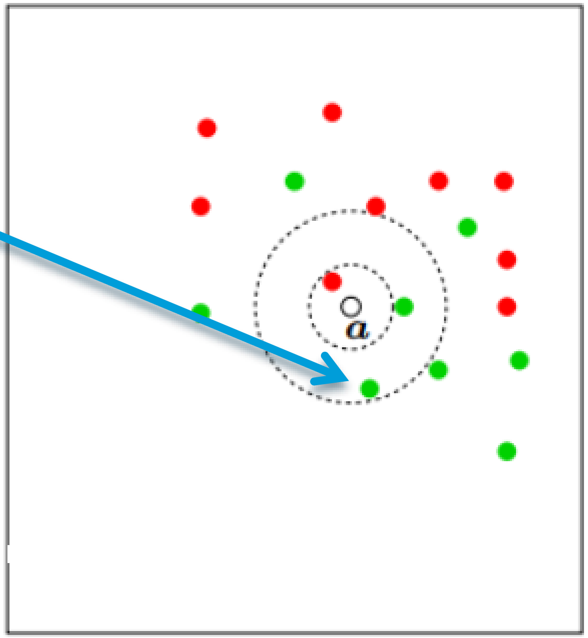
$$euclidean\ distance = \sqrt{(a' - a)^2 + (b' - b)^2}$$



- Nearest neighbour is red classify ‘a’ as red
- For K=1



- 2 out of 3 nearest neighbours are green classify a as green
- For K=3



- We always consider Odd instances in Euclidean circle.

SVM Classifier

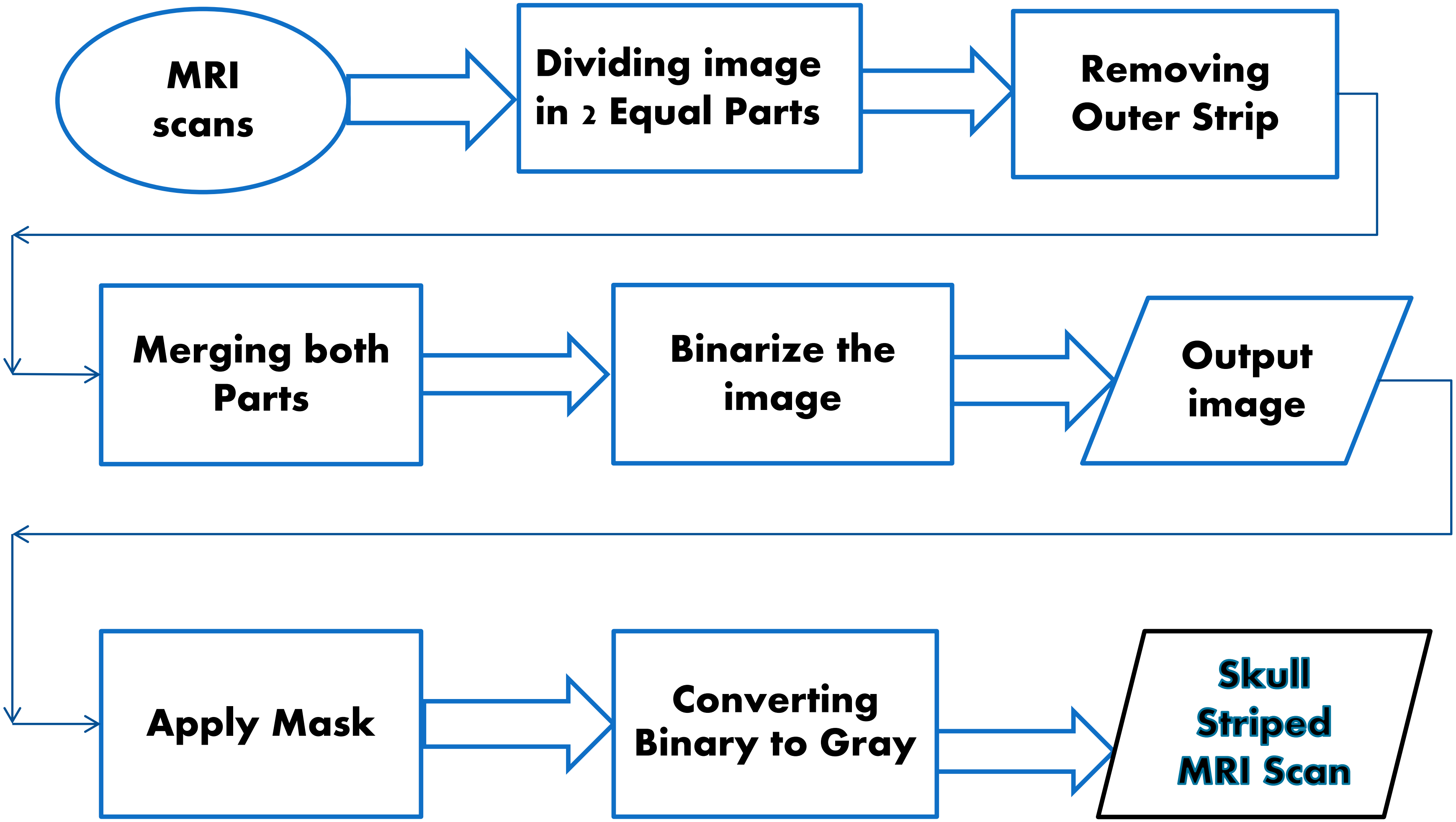
- A supervised classifier.
- Maps features to a higher dimensional space .
- Learns a maximum-margin hyper-plane from labeled training samples of different classes.
- Hyper-plane serves as a boundary between classes.
- Kernels:
  - Linear
  - Polynomial

Proposed Approach

Why Preprocessing ?

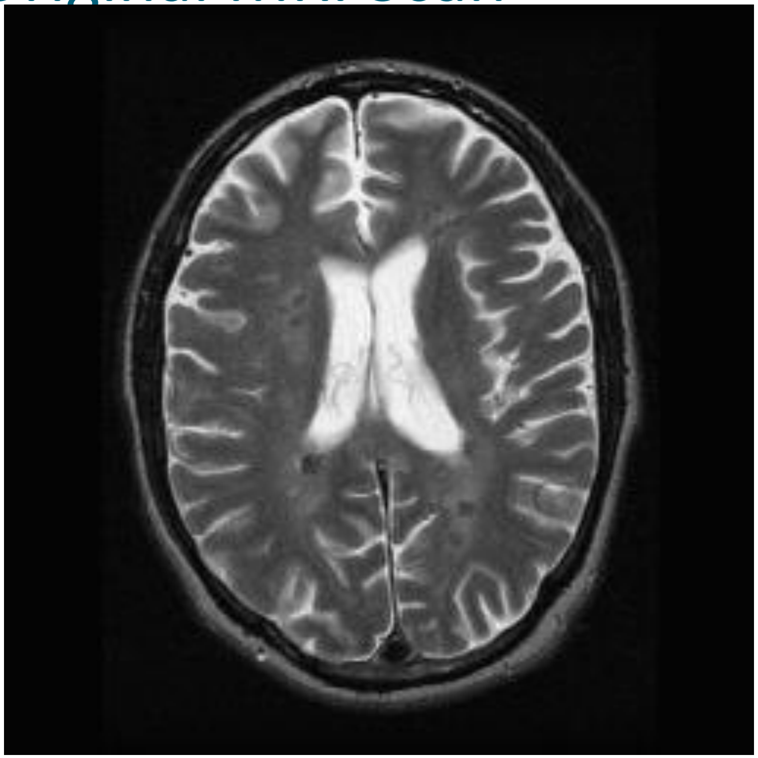
- Skull removing

Pre Processing Flow Chart

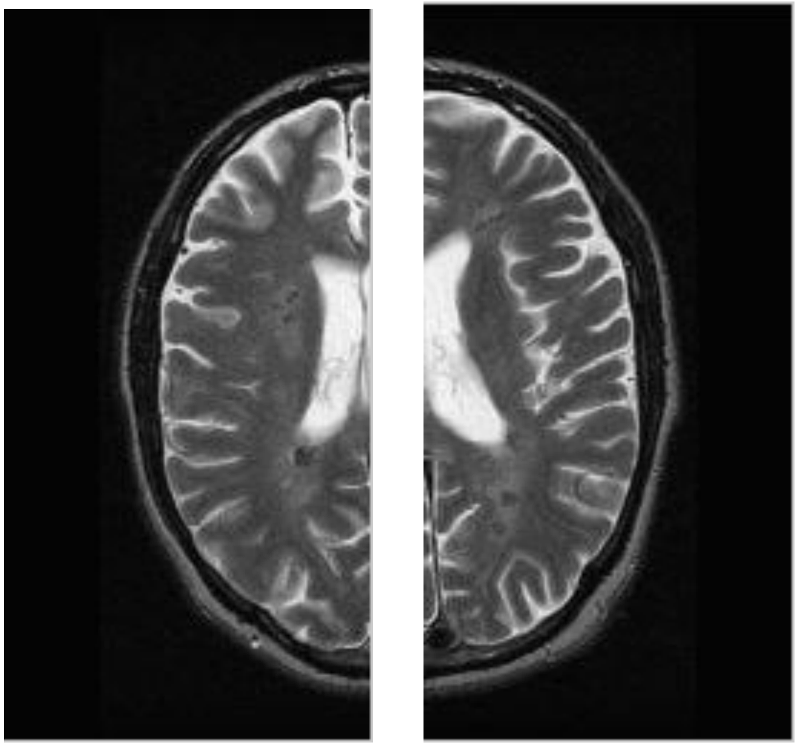


Step 1:

Original MRI Scan

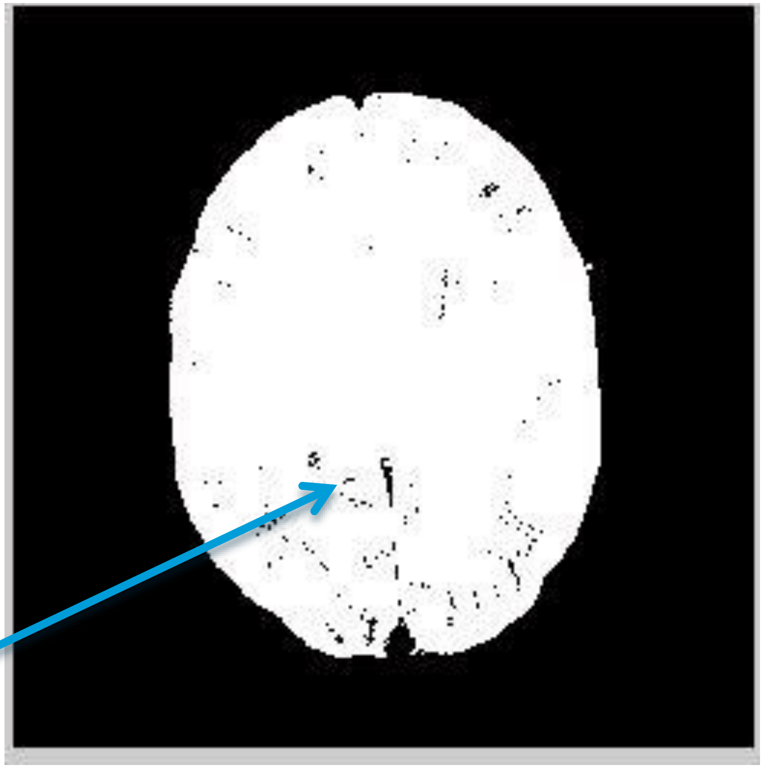


After Division



Step 2,3

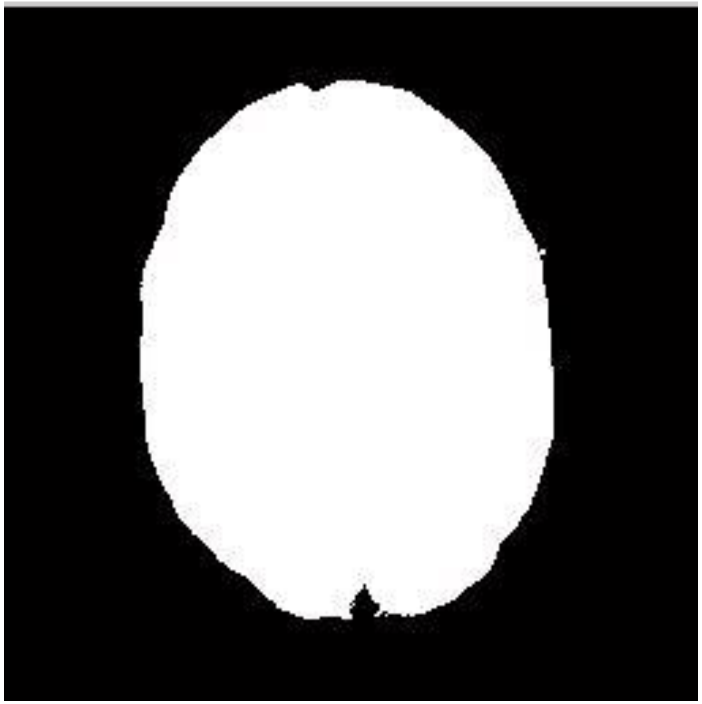
After Removing Skull (Merged)



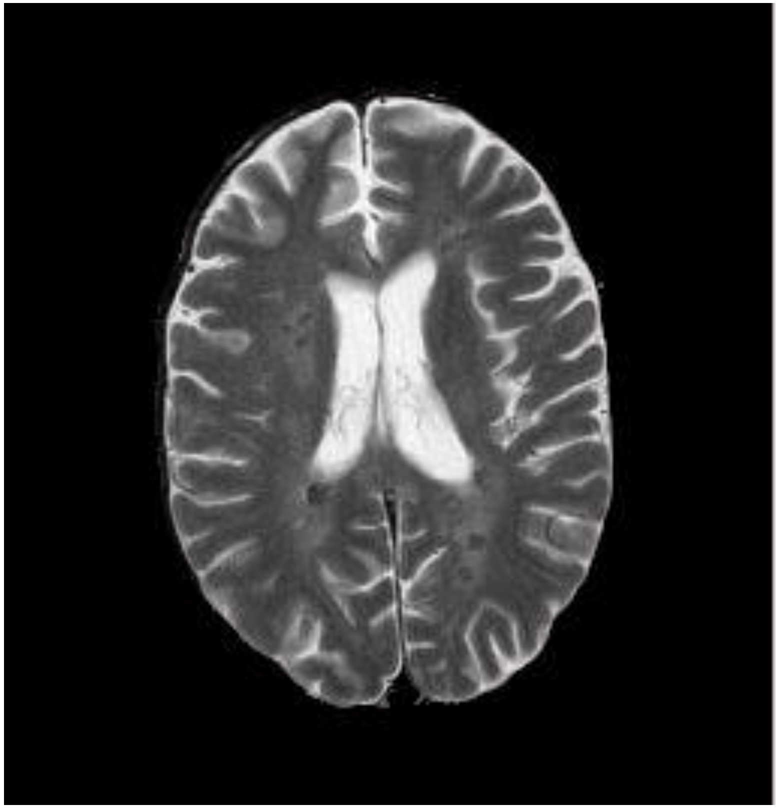
HOLES

Step 4,5&6

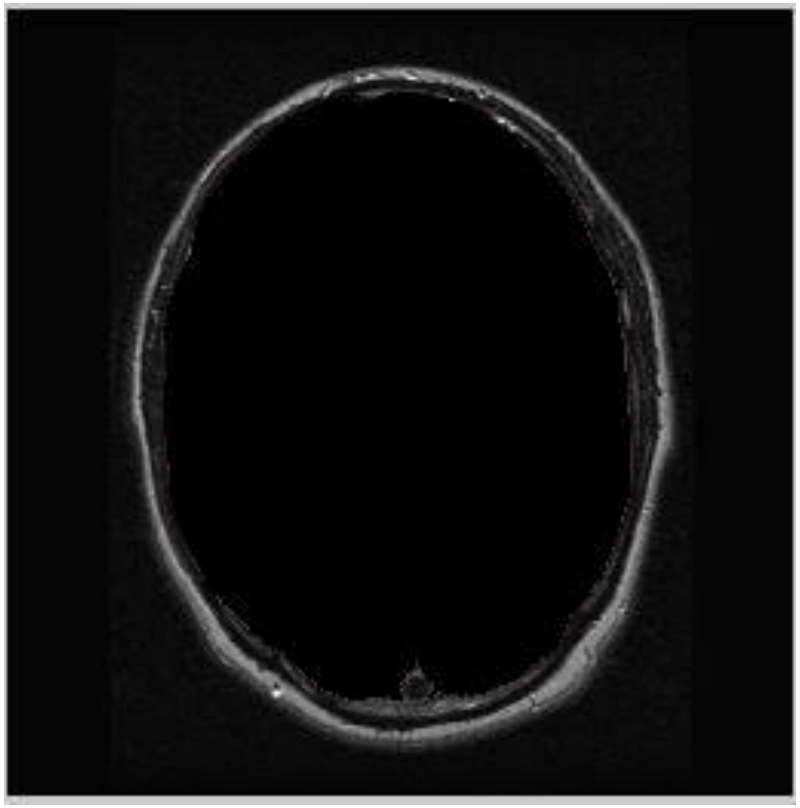
After Removing Holes



Product of Original & B/W

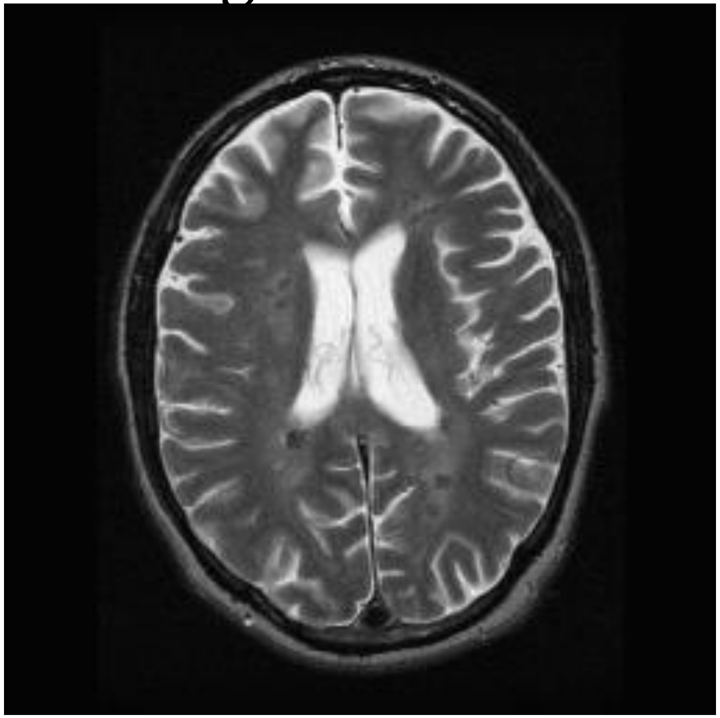


Stripped Part

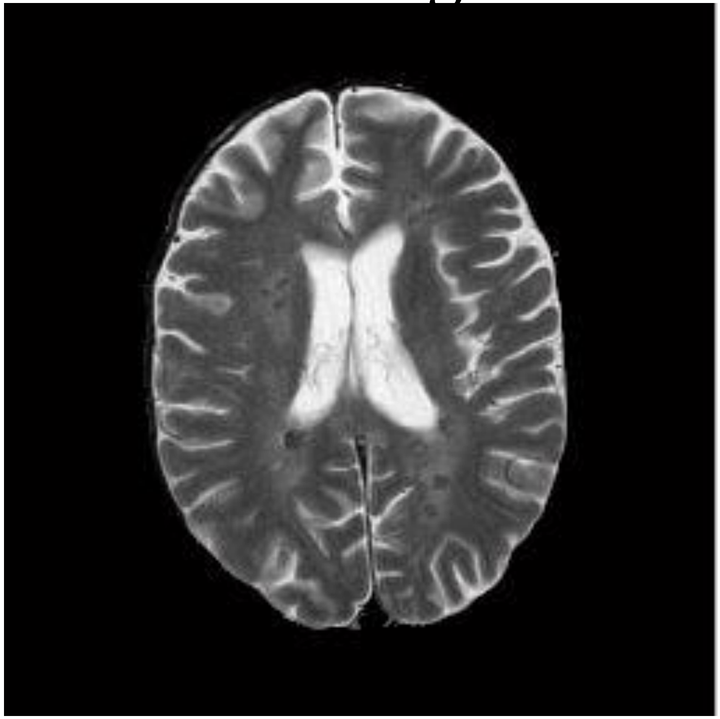


Comparison

Original



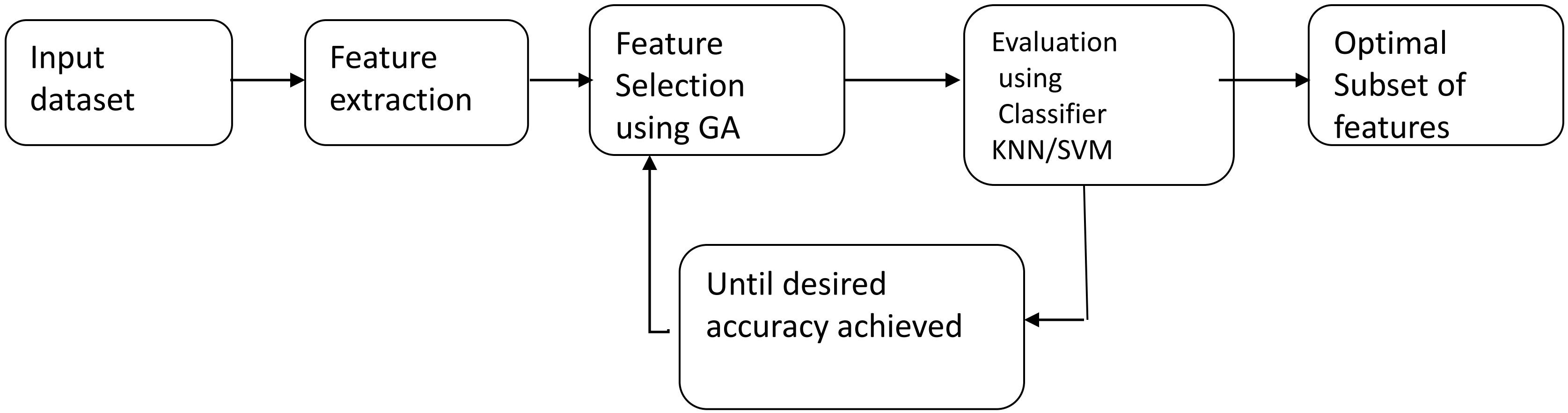
After Processing



Proposed Approach

- Hybrid Machine Learning and Swarm Intelligence
  - Combine different texture feature calculation methods
  - Apply Swarm Intelligence to select features
  - Use selected features in a Machine Learning setup for classification
  - Propose a hybrid classifier with optimal efficiency and performance

Proposed Approach



Over all system

Classifiers used

- Two Different Classifiers are evaluated

  1). K Nearest Neighbors (KNN)

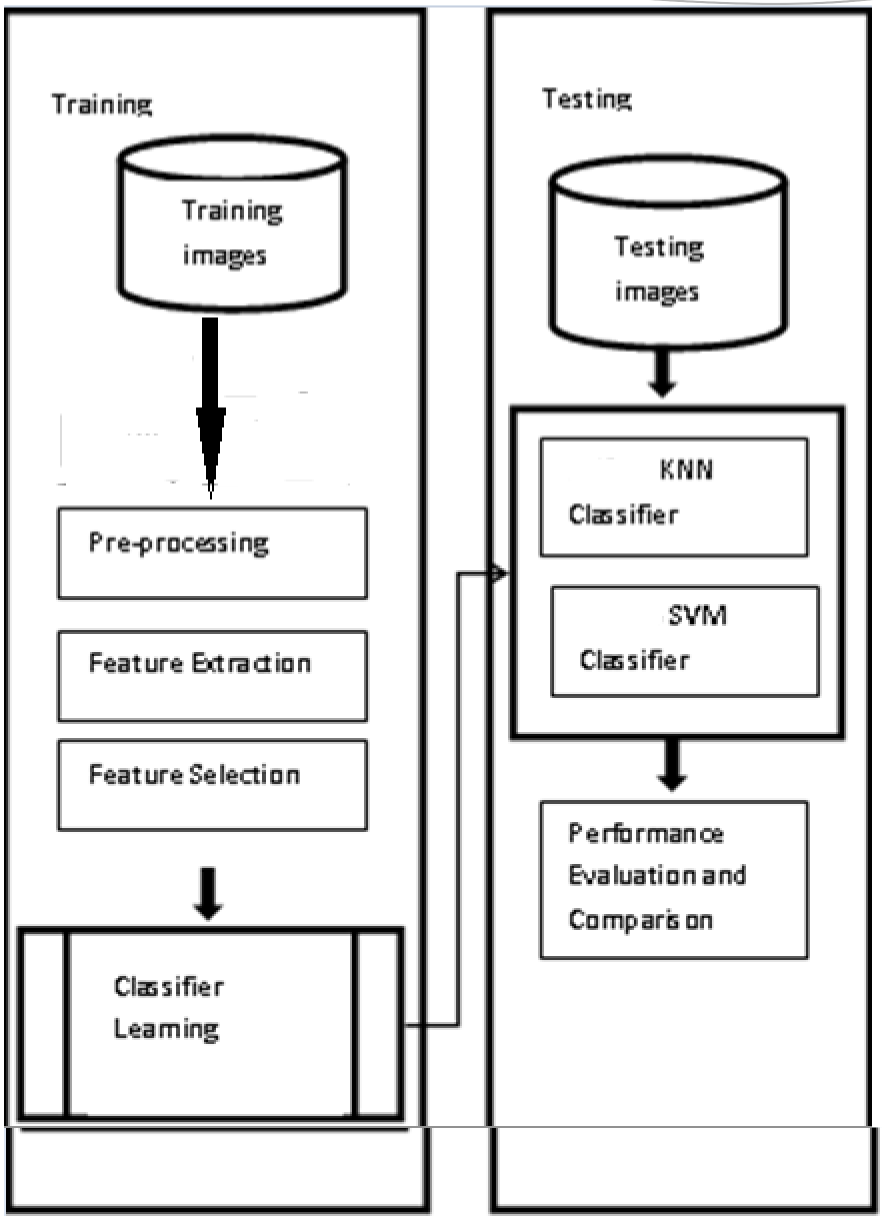
  2). Support Vector Machine (SVM)

10 Fold validation

- Each data set is divided in 10 equal parts.
- 9 parts are selected for training.
- One part is used for testing.
- The process is repeated 10 times to compute mean accuracy.

EXPERIMENTAL RESULTS

System Setup

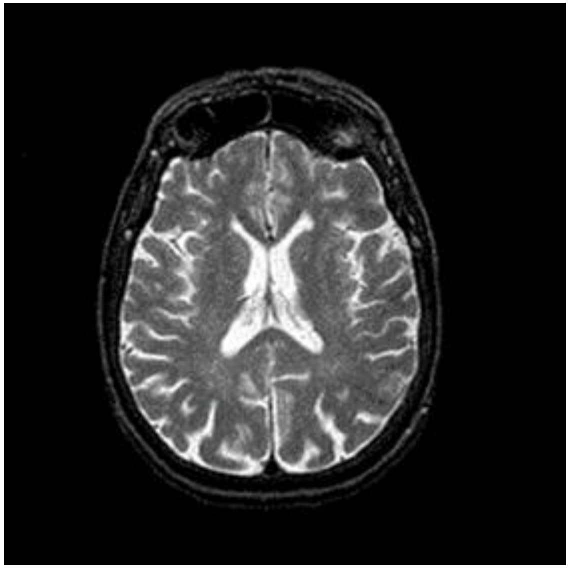


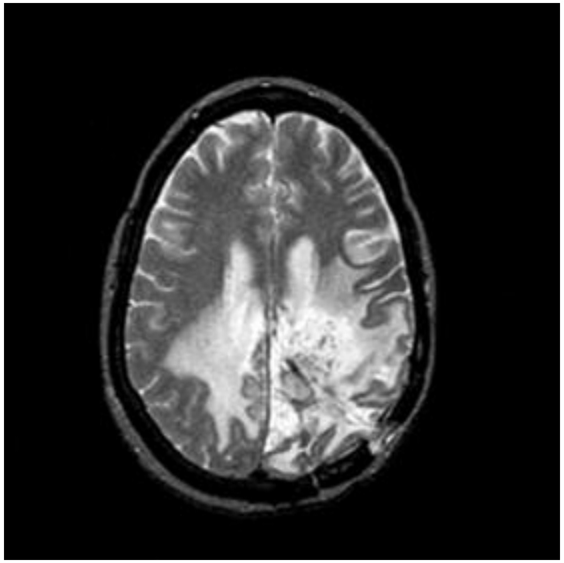
Datasets

Dataset 1). Harvard Medical School (60 image samples, size 256 x 256).(1)

(1)www.med.harvard.edu/AANLIV/

Sample images Dataset



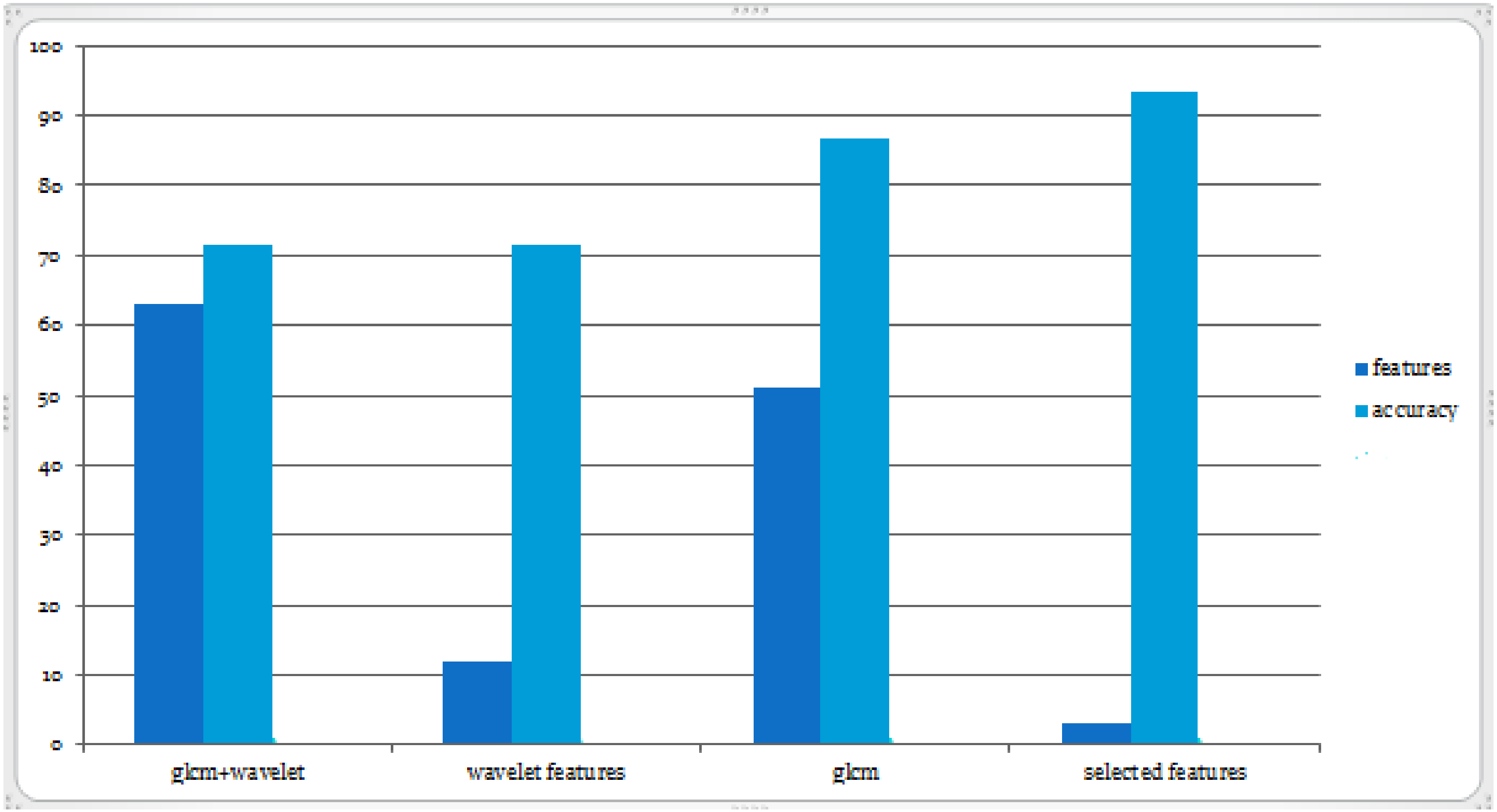


Results

Results GA-KNN K=1,K=3, K=5 Harvard Dataset

Sr. No.	Accuracy	Number of features	Value of K	Features
1	86.6 %	2	K=3	Diff_entropy(r), Variance(v)
2	93.34 %	3	K=1	Diff_entropy(m), Mean_y(m), Deviation_x(r)
3	96.67%	2	K=5	Variance(r), entropy(r)
4	70%	12	K=5	All wavelet features.
5	75%	51	K=5	All GLCM features.
6	70%	63	K=5	All features (GLCM + Wavelet).
7	66.67%	12	K=3	All wavelet features.
8	76.67 %	51	K=3	All GLCM features.
9	66.7 %	63	K=3	All features (GLCM + Wavelet).
10	71.66 %	12	K=1	All wavelet features.
11	86.67 %	51	K=1	All GLCM features.
12	71.67 %	63	K=1	All features (GLCM + Wavelet).

GRAPH FOR K=1



100
90
80
70
60
50
40
30
20
10
0
features
accuracy
gkm+wavelet
wavelet features
gkm
selected features

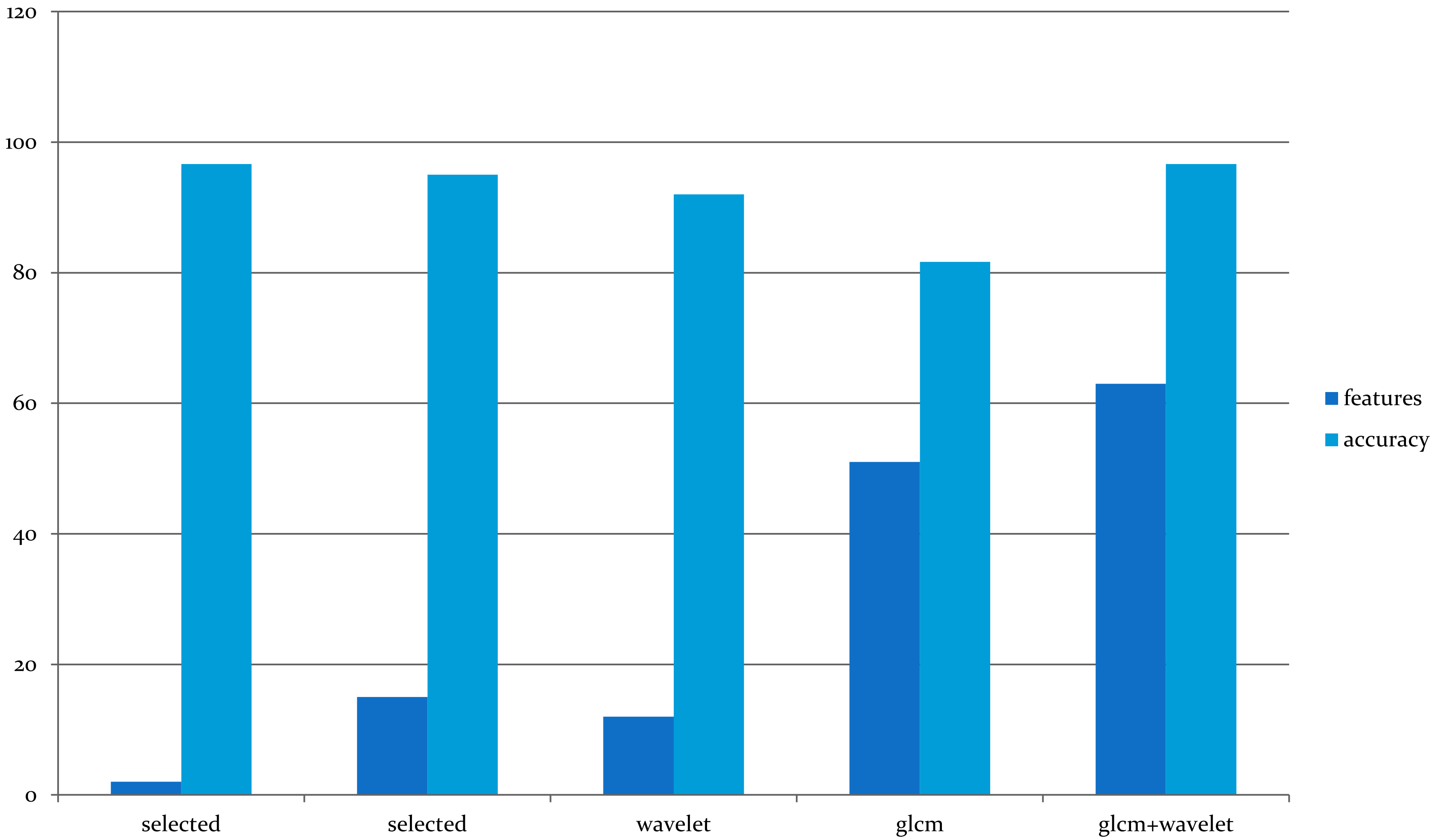
Confussion matrix(3 feature set,KNN,K=1)

	Normal	Abnormal	
Normal	**18**	**2**	**90%**
Abnormal	2	38	95%
	90%	95%	93.3%

Results GA-SVM Polynomial degree-3

Sr. No.	Efficiency	Number of features	Features
1	96.67%	2	Homogeneity(m),sum_average(m)
2	95%	15	Diff_entropy(m), contrast(m),mean_y(m),diss similarity(r), homogeneity(r) ,maximum probability, inertia(r),difference entropy(r), sum average(r),variance(r),entropy(r) ,contrast(r),energy(r),homogeneity(v),energy(v)
3	92%	51	All GLCM features
4	81.67	12	All Wavelet Transform Features.
5	96.67	63	All Features (GLCM+Wavelet)

FOR POLYDEGREE 3



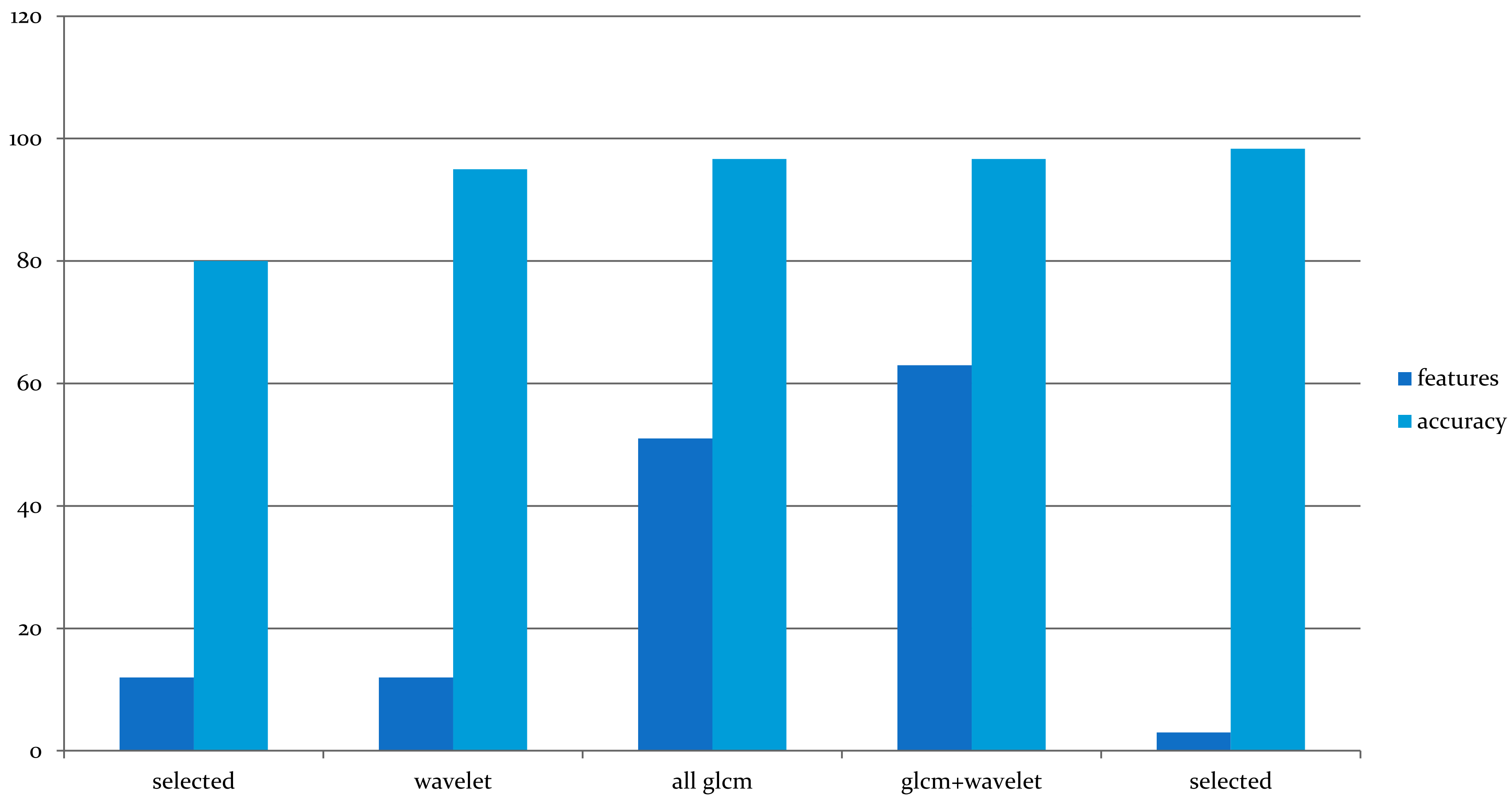
Feature 63 SVM, polynomial degree 3

	Normal	Abnormal	
Normal	20	0	100%
Abnormal	2	38	95%
	90.9%	100%	96.67%

Results GA-SVM Linear

Sr. No.	Efficiency	Number of features	Features
1	98.34%	3	Difference entropy(m),mean_y(m), diviation_x(r)
2	95%	12	Shape(m), Difference entropy(m),contrast(m),diviation_y(m), mean_y(m), homogeneity (r),diviation_x(r) ,sum entropy(v)variance(v), mean(sym2_1D),varience 4(db5_1D)
3	96.67%	51	All GLCM features
4	80%	12	All Wavelet Transform Features.
5	96.67	63	All Features (GLCM+Wavelet)

SVM-LINEAR



120
100
80
60
40
20
0
selected
wavelet
all glcm
glcm+wavelet
selected
features
accuracy

Confussion matrix(3 feature set linear)

	Normal	Abnormal	
Normal	**20**	**0**	**100%**
Abnormal	1	39	97.5
	95.23%	100%	98.34%

Best Results Summary

Sr. No.	Accuracy	Number of features	SVM Kernel
1	98.34 %	3	SVM-Linear
2	93.34 %	3	K= 1
3	86.67 %	2	K=3
4	96.67 %	2	K=5
5	96.67 %	2	SVM-Polynomial Degree 3