Application Of Supervised Machine Learning To Characterize Brain Tissue And To Discriminate Benign Lesions, Various Grades Of Glioma And Metastasis

Tapan Kr Biswas [1*], Anindya Ganguly[2], Rajib Bandopadhyay[1], Ajoy Kr. Dutta[3]

1=Department of Instrumentation and Electronics Engineering, Jadavpur university, India.
2= College of Health and Human Sciences, Charles Darwin University, Australia,
3= Department of Production Engineering, Jadavpur University, India.

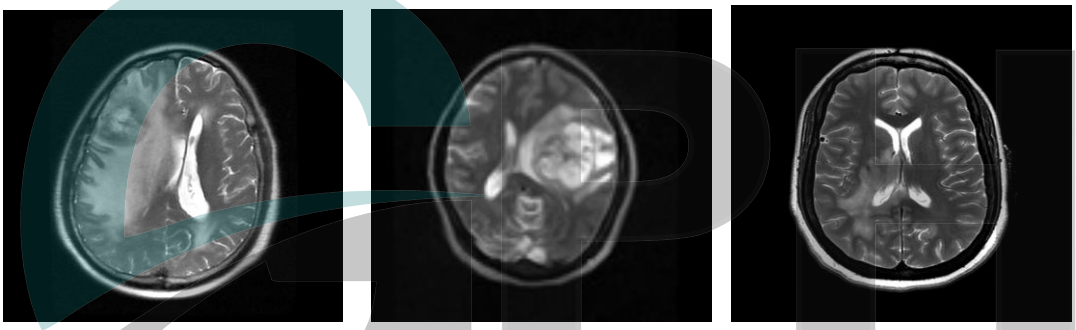
*= Corresponding Author- Email-tbiswas52@gmail.com

ABSTRACT: *Supervised Machine Learning (SML) an extremely powerful classifier was applied for diagnosing the various pathological lesions in the brain, like edema, multiple sclerosis MS), glioma of different grades and metastasis. MR Images may show structural changes in the brain lesions (Figure 1). MR Spectroscopy can also show change in the metabolite peaks and quantities in different disease state (Figure 2). But it is frequently difficult to diagnose the exact disease. Use of SML by two strategies like Artificial Neural Network (ANN) and Support Vector Machine (SVM) helps identifying the condition in doubtful cases. The SVM and ANN train on data sets gathered from different patients based on input variables – Refractive Index,T2 relaxation values, Choline (CHO), Apparent Diffusion Coefficient (ADC), Creatine (CR), CHO/NAA (N-acetyl aspartate), CR/NAA, LIP/LAC (Lipid/lactate), MI (Myoinositol), CHO/CR and T2 value in the periphery of lesion. Refractive index is a vital physical parameter. After training the data, prediction by ANN and SVM show high accuracy in diagnosis. The training and testing have been carried out by Neural Tool in ANN and SVM classifier tool in MATLAB respectively.*

KEY WORDS: Refractive Index (RI),**MR Spectroscopy, Metabolites, Artificial Neural network (ANN), SVM, Error Correcting Code (ECOC), Classifier, Hyperplane, Brain Lesions.**

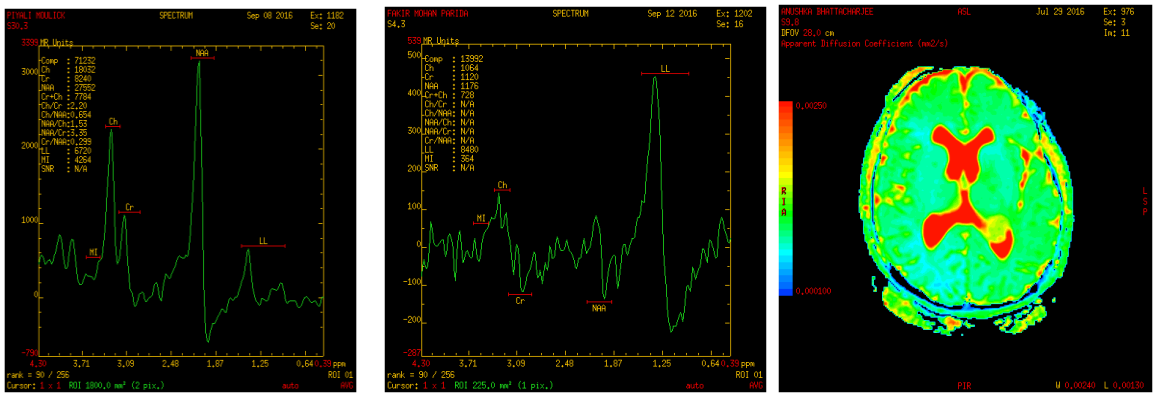
INTRODUCTION: Accurate diagnosis is required for life saving treatment of tumours and different other diseases of brain. Particularly differentiating benign from malignant lesion is mandatory. There is a need for tissue discrimination which is not possible by noting the morbid changes from the MR images only (Figure 1). Images of malignant lesions such as Glioma in

different stages, metastasis, lymphoma and benign diseases like abscess, multiple sclerosis (tumefactive or relapsing remitting Multiple Sclerosis) create confusion sometimes (1). Live prediction of diseases and of the tissue is plausible by data analyzing method of supervised machine learning (2). Two well accepted strategies of supervised machine learning and classification like Artificial Neural Networks (ANN) and Support Vector Machines (SVM) have been implemented separately (3). From the previous research work of the authors (4,5,6) physical data like Refractive indices of tissues and tumours, T2 relaxation and Apparent Diffusion Coefficient (ADC) values from the MRI and different chemical metabolites available from the MR Spectroscopy (MRS) like N Acetyl Aspertate (NAA), Choline (CHO),Creatine(CR), Lipid (Li), Lactate (La) Myoinisitol (MI) and ratio of these metabolites have been collected (7, 8).



A B C

Figure 2. MRI of A. Glioma B.Glioblastoma C.Tumefactive MS



A B C

Figure2. A AND B. MR Spectroscopy showing metabolites and C. ADC mapping to get ADC value

Background of prediction of diseases by using Supervised Machine Learning:

Over the last twenty years Supervised Machine Learning has turned out to be the foundation of information technology. Due to the outpouring of innumerable data there is justification to consider that thorough data analysis will be even more purposeful for technological advancement. Underlying science of machine learning is to resolve the issues and offers good promise for the solutions (9). Artificial ANN and SVM have been executed to increase the diagnostic accuracy in MR examination (MRI and MR Spectroscopy) of the patients suffering from different diseases comparable with the biopsy or histo-pathological study albeit their working procedure is different(6,10) .

A N N: In this study a nonlinear modeling Probabilistic Neural Network technique (PNN) was implemented to evaluate and make virtual pathological prediction from the data obtained from MRI, various metabolic components and their ratio from MRS as mentioned above (Table 1) (11). Application of ANN can live predict the diseases 90 to 95% accurately. The ANN with nonlinearity and extraordinary data processing uniqueness, with generalization capability may be used to characterize the tissue. Thus there are 10 input nodes or independent numeric variables. These variables were placed as column and rows in the spreadsheet and the dependent variables

(diseases and tissues) which were to be predicted were kept in the extreme left of the column. The network consists of a single hidden layer with 10 (multiple) nodes (10). It has several output nodes (such as 8) as well representing different types of tissue (such as CSF, Gray and white matters) and diseases (or types of lesions) (11). These diseases (pathological condition) were edema, cyst, Multiple Sclerosis (MS), low and high grade glioma, lymphoma and metastasis.

SVM :

In this study a nonlinear classification method (12,13) was utilized by SVM. It was applied to evaluate and predict virtual pathological lesions using the data obtained from MRI, metabolic components of MRS (Table 5) and (Figure1, 2). Extraordinary data processing ability of SVM

with nonlinearity and learning aptitude was applied to characterize the disease. Thus there are 11 independent numeric variables. SVM and ANN both belong to supervised learning methods, but their working procedure is different (9,14).

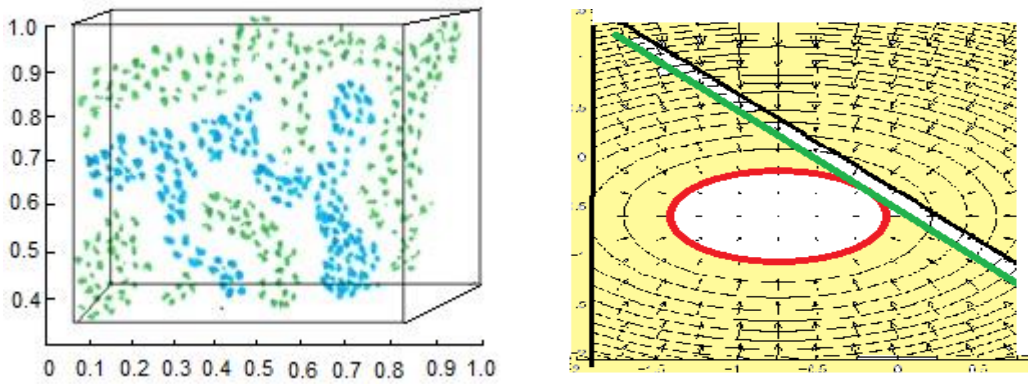
To diminish the errors in classification Error Correcting Output Codes (ECOC) classifier was implemented. ECOC is a special type classifier used to get multiclass learning by reducing multiple binary classifiers (13,14). To train the data a Classification ECOC classifier "fitcecoc" function of the Statistics Toolbox was used.

Self-determination of binary classifiers is the major characteristic for the accuracy or success of performance of ECOC methods (13).

A coding design is necessary to work effectively in multiple classes. A training of the Binary learners and a decoding system determined the prediction of the binary classes and to create a yes/no answer assigned in a set of observations.

i) The design or plan of coding will be one-versus-one in data of two classes or actually "One-against-all method "in multiple numbers of classes. It constructs k SVM models where k is the number of classes. (ii) SVM becomes the learner, (iii) Loss "g" would be utilized by decoding procedure. "g" or Gamma is the parameter of a Gaussian Kernel to handle non-linear classification (Figure 3).

Flattened paraboloid *f:* $2x2+2y2=0$ *with* superimposed constraint *g:* $x+y = 1$. *Minimize* when the constraint line *g* (shown in green in Figure 3 B) is tangent to the inner ellipse contour line of *f* (shown in red)



A **B**

Figure 3. SVMs with two and multiple classes of hyperplanes.

SVM utilizes a hyper plane or function (12) in the midst of various groups of variables or classes to separate them in such a way that each cluster or group persists on both side of the plane showing a particular margin.

SVM twiddles with one dependent and 11 independent numerical variables and 8 output results or groups of different types of brain lesions like MS, different grades of glioma, metastasis and tissue like gray and white matter, CSF. In contrast to other multiclass model ECOC mode can produce enhanced classification and accuracy.

METHODS:

DATA COLLECTION : After taking proper Institutional ethics 137 patients of all ages and genders were examined. RI of different tissues of biopsy materials like gray and white matter, CSF, Glioma of different stages including Glioblastoma, MS (tumefactive), metastasis from lung and breast Ca were determined by Abbe Refractometer. The T2 values were determined by T2 mapping using an Echo Train and various metabolites like Cho, NAA, Cr, MI, Lipid and lactate were noted from the MRS using different TE (35 to 144 ms) and TR -200ms applying PRESS. ADC values were determined by ADC mapping in the MR scanner. A 3 Tesla MR Scanner (GE HDXT, USA) was used to get these data (5).

A N N : For live prediction in ANN, Neural Tool 7.5 (Palisade INC) was used to analyze the data (Table 1) obtained from the previous research work of the author (5,6).

TABLE 1: Tissues, metabolites and ADC,RI,T2 values

Tissue	ADC	CHO	CR	CH/CR	CHO/NAA	CR/NAA	LIP/LAC	MI	RI	T2 (ms)
CSF	300	1610	1400	1.15	0.402	0.346	1400	910	1.3333	400
CSF	320	1680	1800	1.14	0.412	0.367	1760	1056	1.3334	395
CSF	330	1700	1967	1.15	0.432	0.389	1600	1076	1.3335	390
CSF	340	1890	1989	1.14	0.498	0.411	1675	1080	1.3336	388
MS	145	11750	8320	1.4	0.779	0.557	4160	2912	1.3421	340
MS	135	8904	2800	3.15	1.39	0.433	4490	5576	1.3437	328
MS	124	7896	4560	1.73	0.389	0.225	3570	3536	1.3481	316
MS	120	5947	5400	1.1	0.873	0.7396	6766	4294	1.3491	304
MS	75	3448	3320	1.02	0.821	0.7112	5423	2322	1.3588	230
MS	73	1610	2212	0.495	0.465	0.941	1440	364	1.3612	226
g.matter	76	1601	2209	0.491	0.461	0.938	1441	362	1.3956	176
g.matter	76	1601	2209	0.491	0.461	0.938	1441	362	1.3956	176

g.matter	**78**	**1589**	**2219**	**0.491**	**0.459**	**0.941**	**1467**	**345**	**1.3957**	**177**
gmatter	**80**	**1458**	**2320**	**0.494**	**0.456**	**0.878**	**1443**	**321**	**1.3952**	**175**
w matter	**70**	**1180**	**2443**	**0.488**	**0.453**	**0.788**	**1345**	**312**	**1.4251**	**85**
w matter	**71**	**1108**	**2435**	**0.468**	**0.447**	**0.771**	**1341**	**320**	**1.4256**	**83**
matter	**77**	**1098**	**2387**	**0.467**	**0.445**	**0.774**	**1211**	**321**	**1.4259**	**81**
cyst	**84**	**1231**	**2216**	**0.467**	**0.443**	**0.776**	**1123**	**325**	**1.3741**	**193**
cyst	**130**	**1331**	**2321**	**0.456**	**0.442**	**0.787**	**1011**	**321**	**1.3823**	**182**
cyst	**128**	**1298**	**2314**	**0.454**	**0.441**	**0.781**	**1009**	**314**	**1.3821**	**182**
cyst	**131**	**1444**	**2310**	**0.445**	**0.441**	**0.778**	**1001**	**313**	**1.3822**	**184**
glioma	**127**	**1443**	**2243**	**0.423**	**0.431**	**0.766**	**989**	**310**	**1.4331**	**51**
glioma	**177**	**1365**	**2254**	**0.343**	**0.341**	**0.712**	**917**	**300**	**1.4446**	**41**
glioblst	**156**	**2655**	**2112**	**0.311**	**0.332**	**0.678**	**900**	**311**	**1.4551**	**38**
glioblst	**142**	**2774**	**3280**	**0.844**	**0.907**	**1.06**	**2240**	**312**	**1.4512**	**36**

Independent variables were considered as 10 inputs (10) **:**

RI values

T2 value

ADC value

Quantities of metabolites (Choline,Creatine, MI ,NAAaa, lipid/ lactate)

Ratio of Choline, NAA

Ratio of Creatine, NAA

To live predict (decision)

Diseases or tissues are regarded as dependent variables

ANN (10,11) :

1. In the excel spread sheet the data were tabulated (Table.1) and the Dependent Variable(Disease or tissues) remained in the extreme left column of the data table and Independent Numeric variable (Usually RI value, ratio of Choline NAA or ADC value) in the corresponding columns in the right side of the table.

2. A **data set manager** was prepared from the values of the excel sheet.
3. Training and testing of the values of table 2 were done statistically and mentioned in Table 4.

TABLE 2. Testing and training of the variables

ADC	CHO	CR	CH/CR	CHO/NAA	CR/NAA	LIP/LAC	MI	RI	T2	DISEASE
300	1610	1400	1.15	0.402	0.346	1400	910	1.3333	400	CSF
320	1680	1800	1.14	0.412	0.367	1760	1056	1.3334	395	CSF
330	1700	1967	1.15	0.432	0.389	1600	1076	1.3335	390	CSF
340	1890	1989	1.14	0.498	0.411	1675	1080	1.3336	388	CSF
145	11750	8320	1.4	0.779	0.557	4160	2912	1.3421	340	MS
135	8904	2800	3.15	1.39	0.433	4490	5576	1.3437	328	MS
124	7896	4560	1.73	0.389	0.225	3570	3536	1.3481	316	MS
120	5947	5400	1.1	0.873	0.7396	6766	4294	1.3491	304	MS
75	3448	3320	1.02	0.821	0.7112	5423	2322	1.3588	230	MS
73	1610	2212	0.495	0.465	0.941	1440	364	1.3612	226	MS
76	1601	2209	0.491	0.461	0.938	1441	362	1.3956	176	gmatter
76	1601	2209	0.491	0.461	0.938	1441	362	1.3956	176	gmatter
78	1589	2219	0.491	0.459	0.941	1467	345	1.3957	177	gmatter
80	1458	2320	0.494	0.456	0.878	1443	321	1.3952	175	gmatter
70	1180	2443	0.488	0.453	0.788	1345	312	1.4251	85	w matter
71	1108	2435	0.468	0.447	0.771	1341	320	1.4256	83	w matter
77	1098	2387	0.467	0.445	0.774	1211	321	1.4259	81	w matter
84	1231	2216	0.467	0.443	0.776	1123	325	1.3741	193	cyst
130	1331	2321	0.456	0.442	0.787	1011	321	1.3823	182	cyst
128	1298	2314	0.454	0.441	0.781	1009	314	1.3821	182	cyst
131	1444	2310	0.445	0.441	0.778	1001	313	1.3822	184	cyst
127	1443	2243	0.423	0.431	0.766	989	310	1.4331	51	glioma
177	1365	2254	0.343	0.341	0.712	917	300	1.4446	41	glioma
156	2655	2112	0.311	0.332	0.678	900	311	1.4551	38	glioblst
142	2774	3280	0.844	0.907	1.06	2240	312	1.4512	36	glioblst

Train-Test Report for Net Trained on Data Set #1

Tag Used	Prediction	Prediction%	Incorrect%	Good/Bad
train				
train				
train				
train				
train				
test	CSF	20.00%	80.00%	Bad
test	CSF	20.00%	80.00%	Bad
train				
train				
train				
test	gmatter	100.00%	0.00%	Good
train				
train				
train				
train				
test	w matter	100.00%	0.00%	Good
train				
train				
train				
test	cyst	100.00%	0.00%	Good
train				
train				
train				
train				
train				

Table 3. Variables and Prediction

RI	T2	L/L	NAA	CR	CHO	MI	DISEASE	Tag Used	Prediction	P
1.3333	400	3	13	7	6.8	4	CSF			
1.3334	395	3	13	7	6.8	4	CSF			
1.3335	390	3	13	7	6.8	4	CSF			
1.3336	388	3	13	7	6.8	4	CSF			
1.3337	385	3.3	13	7	6.8	3.55	EDEMA			
1.3338	381	3.3	13	7	6.8	3.55	EDEMA			
13339	378	3.3	13	7	6.8	3.55	EDEMA			
1.334	375	3.3	13	7	6.8	3.55	EDEMA			
1.335	371	3.3	13	7	6.8	3.55	EDEMA			
1.3361	368	3.3	13	7	6.8	3.5	EDEMA			
1.3381	364	3.3	13	6.9	6.8	3.5	EDEMA			
1.3391	361	3.3	13	6.9	6.8	3.5	EDEMA			
1.3411	358	3.3	13	6.9	6.8	3.5	EDEMA			
1.3412	355	3.3	13	6.9	6.8	3.5	EDEMA			
1.3413	351	3.3	13	6.9	6.8	3.5	EDEMA			
1.3416	348	3.3	13	6.9	6.8	3.5	EDEMA			
1.3418	345	3.3	13	6.9	6.8	3.5	EDEMA			
1.3419	342	3.3	13	6.9	6.8	3.5	EDEMA			
1.3421	340	3.3	11	6.8	6.77	3.45	MS			
1.3437	328	3.4	10	6.8	6.77	3.45	MS			
1.3481	316	3.4	10	6.8	6.77	3.45	MS			
1.3491	304	3.4	10	6.8	6.77	3.45	MS			
1.3511	292	3.4	10	6.8	6.76	3.45	MS			
1.3521	280	3.4	10	6.8	6.76	3.45	MS			
1.3531	268	3.4	9	6.8	6.76	3.45	MS			
1.3533	256	3.4	9	6.8	6.76	3.45	MS			
1.3534	244	3.5	9	6.8	6.76	3.44	MS			
1.3533	236	3.5	8	6.8	6.76	3.44	MS			
1.3588	230	3.5	8	6.8	6.76	3.44	MS			
1.3612	226	3.5	8	6.8	6.76	3.44	MS			
1.3791	220	4	14	8	7	3.5	CYST			
1.3811	211	4	14	8	7	3.5	CYST			
1.3812	202	4	14	8	7	3.5	CYST			
1.3843	193	4	14	8	7	3.5	CYST			
1.3925	182	4	14	8	7	3.5	CYST			
1.3952	175	3	15	7	6.8	3.55	G Matter			
1.3978	160	3	15	7	6.8	3.55	G Matter			

1.4012	140	3	15	7	6.8	3.55	G Matter			
1.4113	120	3	15	7	6.8	3.55	G Matter			
1.4123	100	3	14	7	6.8	3.52	Wmtter			
1.4144	85	3	14	7	6.8	3.52	Wmtter			
1.4169	70	3	13	7	6.8	3.52	Wmtter			
1.4251	60	3	13	7	6.8	3.52	Wmtter			
1.4288	51	2.8	4.8	3	6	3.51	Giloma			
1.4291	51	2.8	4.8	3	6	3.51			predict	Giloma
1.4311	51	2.8	4.8	3	6	3.51			predict	Giloma
1.4315	51	2.8	4.8	3	6	3.51			predict	Giloma
1.4321	51	2.8	4.8	3	6	3.51			predict	Giloma
1.4435	45	1	3	3.2	10	1			predict	Glioblastoma
1.4439	41	1	3	3.2	10	1			predict	Glioblastoma
1.4446	38	1	3	3.2	11	1			predict	Glioblastoma
1.4551	36	1	3	3.2	11	1			predict	Glioblastoma
1.4624	34	10	3	1.8	1	1.8			predict	Lymphoma
1.4676	32	10	3	1.8	1	1.8			predict	Lymphoma
1.4782	31	10	3	1.8	1	1.8			predict	Lymphoma
1.4799	29	10	3	1.8	1	1.8			predict	Lymphoma
1.4834	28	12	2	1.8	1	1.8			predict	METS
1.4911	28	12	2	1.8	1	1.8			predict	METS

Table.4 Neural Net Training and auto testing

NeuralTools: Neural Net Training and Auto-Testing

Net: Net Trained on Data Set #1

Summary	
Net Information	
Name	Net Trained on Data Set #1
Configuration	PNN Category Predictor
Location	This Workbook
Independent Category Variables	0
Independent Numeric Variables	10 (RI, CH/CR, CHO, ADC, CR, CHO/NAA, CR/NAA, LIP/LAC, MI, T2)
Dependent Variable	Category Var. (DISEASE)
Training	
Number of Cases	20
Training Time	0:00:00
Number of Trials	107
Reason Stopped	Auto-Stopped
% Bad Predictions	0.0000%
Mean Incorrect Probability	0.0401%
Std. Deviation of Incorrect Prob.	0.1044%
Testing	
Number of Cases	5
% Bad Predictions	0.0000%
Mean Incorrect Probability	21.5009%
Std. Deviation of Incorrect Prob.	29.5248%
Data Set	
Name	Data Set #1

SVM: The data were tabulated from the previous research of the author (5) containing 135 rows of CSF, MS, Gray and white matters, low and high grade glioma (Astrocytoma Gr III/IV - Glioblastoma) and metastasis from 137 patients of different genders and ages (after taking proper institutional ethics). Due to the space constraint selected 53 rows were depicted in the Table 5. Training of all these data was done by SVM. The SVM is then tested on data from 19 patients for prediction of disease or tissues (Table 6).

The method comprises of two steps (Figure 3):

i) To train the SVM applying on available data of the patient as depicted on Table 5 to get a model data set (Table 6).

ii) Testing the model data set of trained SVM along with the unknown dataset (Table7) to characterize, labeling or classify unknown data.

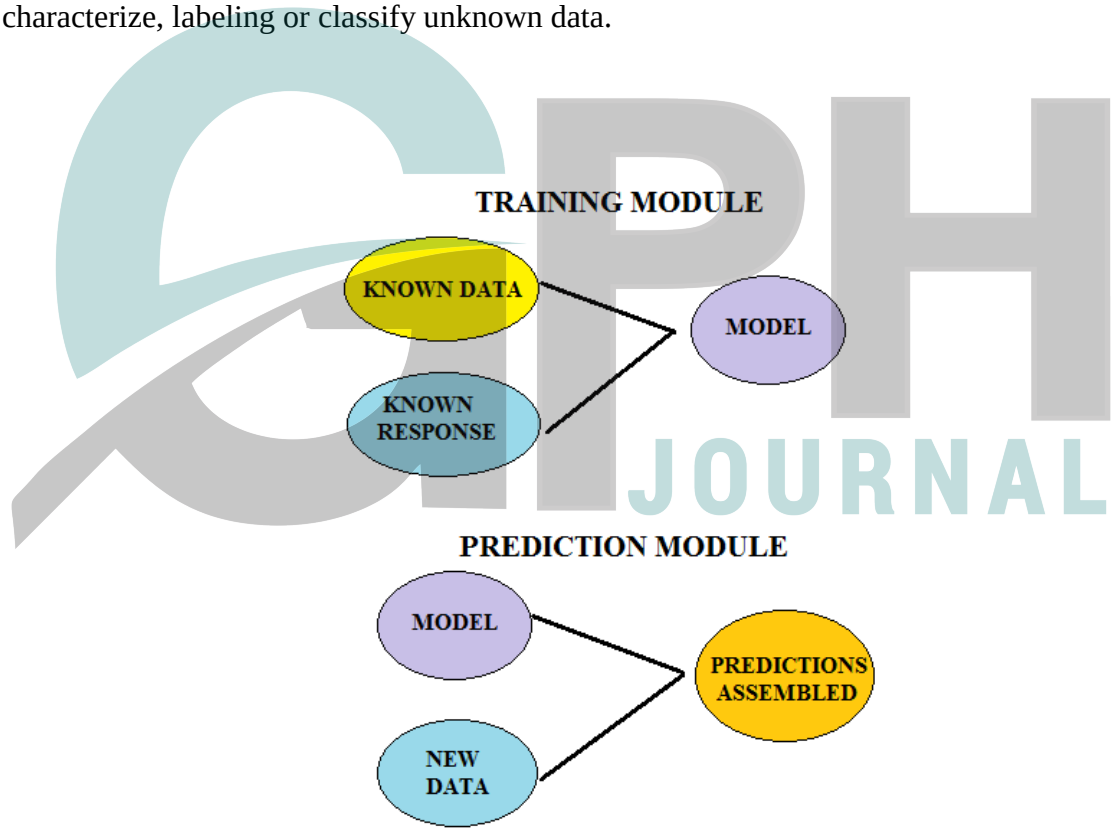


Figure3. Working Plan of SVM

Step 1: The data were identified and labeled at first (Table 5). Label comprises the type of tissue such as CSF, Gray and white matters or diseases like MS, Glioma, metastasis etc. SVM actually correlated the data sets with the appropriate labels. In the spreadsheet, the first column

denotes different types of labels. The other 11 columns consist of the data or numerical variables. As per the Table 5, row No.5 depicts label CSF and that particular CSF scan has 11 values (or independent variables) equivalent to T2, ADC, CR, CHO etc (Figure 4) (15).

Independent variables as inputs:

RI values

T2 value (within 5 mm and 10 mm of outline of the lesion).

ADC value

Quantities of metabolites (CHOLINE,CREATINE,MI ,NAA,LIPID/ LACTATE)

Ratio of CHOLINE: NAA

Ratio of CREATINE : NAA

Step 2: Fitcecoc command was implemented to train the SVM for the data and label using Classification Models (Classification Learner App of 64-bit MATLAB R2017a Environment on windows 10 home platform) (15) SVM used Supervised Learning and classified data. In this case as there are 8 different labels, fitcecoc command was the appropriate to accommodate multiple classes instead of the "FitSVM" or "svmtrain" which is the basic command when only two classes/labels or binary classification are at hand (16)

As a result the variable Md1 happens to be the Support Vector Machine trained using the data (11,12).

Step 3: Using the unknown data sets (Table 7) (which was not used to train the SVM initially). The SVM was tested for prediction. The data in the SVM was run and it predicted the classes or characterized the tissue on its own.

SOURCE CODE AND PLATFORM: To describe, analyze and model the data, Statistics and Machine Learning Toolbox™ of MAT LAB was used for the necessary functions and apps. To perform supervised machine learning a known set of input data (observations) to be provided and known responses to the data (i.e., labels or classes) is noted. To generate code for training and to reconstruct the trained model with new data, or to learn about programmatic classification, the model to be exported to the workspace of tool box.

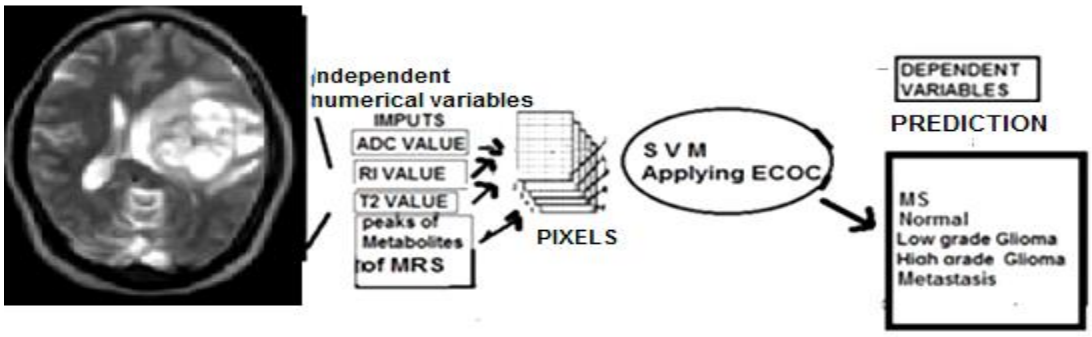


Figure 4. SVM for live prediction of diseases as Dependent variables using independent numerical variables as inputs (Reference 5.)

Table 5. Selected 53 rows from full Data set (Reference 5):

DISEASE	T2	CHO	ADC	CR	CHO/NAA	CR/NAA	LIP/LAC	MI	CH/CR	T2peri	RI
CSF	400	1610	300	1400	0.402	0.346	1400	910	1.15	400	1.3333
CSF	399	1676	307	1450	0.404	0.347	1489	917	1.15	399	1.3333
CSF	398	1689	311	1560	0.408	0.351	1550	957	1.15	399	1.3333
CSF	397	1700	313	1600	0.409	0.357	1554	987	1.15	399	1.3333
CSF	396	1728	320	1788	0.412	0.361	1660	1050	1.14	395	1.3333
CSF	395	1711	322	1800	0.422	0.367	1701	1056	1.14	395	1.3333
CSF	391	1737.0833	330.08333	1974.3056	0.4326389	0.385	1778.1389	1122.1944	1.1405556	390.88889	1.3333
CSF	345	2021	402	2060	0.572	0.448	1744	1145	1.15	389.88889	1.3333
CSF	344	2022	403	2061	0.573	0.451	1744	1145	1.15	388.88889	1.3333
CSF	343	2023	404	2062	0.574	0.452	1745	1146	1.15	387.88889	1.3333
CSF	342	2024	405	2068	0.577	0.453	1746	1147	1.15	386.88889	1.3333
CSF	341	2123	411	2063	0.578	0.453	1747	1148	1.15	385.88889	1.3333
ms	340	11750	145	8320	0.779	0.557	4160	2912	1.4	384.88889	1.3334
ms	339	11750	1460	8319	0.778	0.541	4423	3223	1.4	383.88889	1.3335
ms	338	11749	1459	8314	0.776	0.538	4423	3221	1.4	382.88889	1.3336
ms	337	11746	1445	8311	0.774	0.536	4421	3220	1.4	381.88889	1.3421
ms	336	11745	1444	8310	0.773	0.534	4422	3219	1.4	380.88889	1.3439
ms	335	11745	1443	8309	0.772	0.532	4420	3216	1.4	379.88889	1.3498
ms	334	11743	1443	8308	0.771	0.531	4419	3214	1.4	378.88889	1.3499
ms	333	11742	1442	8306	0.768	0.531	4415	3210	1.4	377.88889	1.35
ms	304	5947	120	5400	0.873	0.7396	6766	4294	1.1	245	1.3501
ms	249	3448	112	3320	0.821	0.7112	5423	2322	1.02	230	1.3589
ms	245	1610	110	2212	0.465	0.941	1440	2276	0.495	227	1.3641
gmatter	130	1601	72	2209	0.464	0.938	1439	361	0.491	166	1.3956
gmatter	129	1599	73	2208	0.463	0.936	1437	357	0.4911	165	1.3956
gmatter	128	1597	74	2206	0.463	0.934	1435	351	0.489	165	1.3956
gmatter	127	1595	75	2204	0.462	0.933	1431	348	0.489	164	1.3956
w matter	95	1180	70	2443	0.453	0.788	1345	312	0.488	148	1.4251
w matter	93	1108	71	2435	0.447	0.771	1341	320	0.468	146	1.4256
w matter	92	1098	77	2387	0.445	0.774	1211	321	0.467	150	1.4259
w matter	91	1006	79	2389	0.445	0.774	1209	322	0.467	156	1.4259
edema	160	1231	132	2216	0.443	0.776	1123	325	0.467	246	1.3741
edema	182	1331	130	2321	0.442	0.787	1011	321	0.456	243	1.3823
edema	182	1298	128	2314	0.441	0.781	1009	314	0.454	244	1.3821
edema	184	1444	131	2310	0.441	0.778	1001	313	0.445	245	1.3822
edema	186	1447	133	2321	0.441	0.778	1000	313	0.445	247	1.3822
edema	187	1449	135	2324	0.441	0.778	1001	313	0.445	246	1.3822
GLIOMA	90	1443	127	2243	0.431	0.766	989	310	0.423	175	1.4331
GLIOMA	99	1365	177	2254	0.341	0.712	917	300	0.343	170	1.4339
GLIOMA	101	1431	179	2259	0.34	0.701	915	300	0.341	181	1.4438
GLIOMA	105	1785	165	2111	0.34	0.701	915	300	0.341	181	1.4446
GLIOMA	105	1812	161	2113	0.34	0.701	912	302	0.339	186	1.4447
GLIOMA	107	2213	155	2114	0.333	0.677	901	310	0.321	191	1.4456
Gblastma	108	2457	154	2115	0.332	0.676	900	311	0.311	195	1.4512
Gblastma	109	2655	152	2112	0.332	0.676	900	311	0.311	195	1.4539
Gblastma	110	2655	144	2912	0.867	0.678	900	311	0.311	195	1.4551
Gblastma	127	1287	133	2596	0.567	0.811	1891	322	0.76654	195	1.4703
METS	129	1298	130	2567	0.511	0.657	1011	323	0.432	200	1.4831
METS	130	1301	130	2478	0.511	0.657	1011	323	0.432	200	1.4831
METS	129	1278	130	2567	0.511	0.657	1011	323	0.432	200	1.4834
METS	133	1311	132	2567	0.511	0.657	1011	324	0.432	200	1.4837
METS	135	1321	135	2532	0.432	0.654	1011	324	0.432	200	1.4845
METS	154	1414	134	2027	0.426	0.715	1122	358	0.454	225	1.4914
METS	155	1415	135	2027	0.427	0.715	1123	359	0.454	226	1.4917

TABLE 6. Training Data set

DISEASE	T2	CHO	ADC	CR	CHO/NAA	CR/NAA	LIP/LAC	MI	CH/CR	T2peri	RI
CSF	400	1610	300	1400	0.402	0.346	1400	910	1.15	400	1.3333
CSF	399	1676	307	1450	0.404	0.347	1489	917	1.15	399	1.3333
CSF	398	1689	311	1560	0.408	0.351	1550	957	1.15	399	1.3333
CSF	394	1710	322	1809	0.423	0.368	1690	1059	1.14	394	1.3333
ms	340	11750	145	8320	0.779	0.557	4160	2912	1.4	393	1.3334
ms	339	11750	1460	8319	0.778	0.541	4423	3223	1.4	392	1.3335
ms	336	11745	1444	8310	0.773	0.534	4422	3219	1.4	391	1.3439
ms	245	1610	110	2212	0.465	0.941	1440	2276	0.495	227	1.3641
gmatter	130	1601	72	2209	0.464	0.938	1439	361	0.491	166	1.3956
gmatter	129	1599	73	2208	0.463	0.936	1437	357	0.4911	165	1.3956
gmatter	128	1597	74	2206	0.463	0.934	1435	351	0.489	165	1.3956
w matter	95	1180	70	2443	0.453	0.788	1345	312	0.488	148	1.4251
w matter	93	1108	71	2435	0.447	0.771	1341	320	0.468	146	1.4256
edema	160	1231	132	2216	0.443	0.776	1123	325	0.467	246	1.3741
edema	191	1451	131	2356	0.441	0.778	990	313	0.445	245	1.3822
edema	193	1452	130	2340	0.441	0.768	990	312	0.445	245	1.3823
GLIOMA	90	1443	127	2243	0.431	0.766	989	310	0.423	175	1.4331
GLIOMA	107	2213	155	2114	0.333	0.677	901	310	0.321	191	1.4456
Gblastma	108	2457	154	2115	0.332	0.676	900	311	0.311	195	1.4512
Gblastma	109	2655	152	2112	0.332	0.676	900	311	0.311	195	1.4539
METS	129	1298	130	2567	0.511	0.657	1011	323	0.432	200	1.4831
METS	130	1301	130	2478	0.511	0.657	1011	323	0.432	200	1.4831
METS	152	1412	132	2022	0.425	0.713	1121	357	0.451	224	1.4913
METS	154	1414	134	2027	0.426	0.715	1122	358	0.454	225	1.4914

First Columns: Labels or class representing tissue/lesions. Other columns corresponding to each class used for training.

Table-7: Test Data Set (Full) Used For Prediction

	A	B	C	D	E	F	G	H	I	J	K	L
1	CSF	347	2017	399	2059	0.569	0.447	1743	1144	1.15	347	1.3333
2	CSF	343	2023	404	2062	0.574	0.452	1745	1146	1.15	343	1.3333
3	ms	326	8876	131	2781	1.38	0.431	4478	5561	1.15	241	1.3507
4	ms	316	7896	124	4560	0.389	0.225	3570	3536	1.73	243	1.3518
5	gmatter	129	1599	73	2208	0.463	0.936	1437	357	0.4911	165	1.3956
6	gmatter	125	1593	77	2200	0.46	0.928	1424	346	0.487	168	1.3956
7	w matter	93	1108	71	2435	0.447	0.771	1341	320	0.468	146	1.4256
8	w matter	89	1012	82	2385	0.444	0.775	1201	324	0.466	165	1.4259
9	edema	182	1331	130	2321	0.442	0.787	1011	321	0.456	243	1.3823
10	edema	187	1449	135	2324	0.441	0.778	1001	313	0.445	246	1.3822
11	GLIOMA	99	1365	177	2254	0.341	0.712	917	300	0.343	170	1.4339
12	GLIOMA	105	1785	165	2111	0.34	0.701	915	300	0.341	181	1.4446
13	Gblastma	109	2655	152	2112	0.332	0.676	900	311	0.311	195	1.4539
14	Gblastma	118	2661	140	3189	0.89	1.02	2134	314	0.7881	192	1.4576
15	Gblastma	128	1284	131	2589	0.541	0.781	1767	322	0.76651	198	1.4723
16	METS	130	1301	130	2478	0.511	0.657	1011	323	0.432	200	1.4831
17	METS	135	1321	135	2532	0.432	0.654	1011	324	0.432	200	1.4845
18	METS	151	1411	131	2019	0.423	0.713	1119	356	0.449	223	1.4911
19	METS	154	1414	134	2027	0.426	0.715	1122	358	0.454	225	1.4914

RESULTS AND DISCUSSION:

A. ANN :

(i) It is evident that the prediction of tissue and diseases was 100% accurate (Table.8) when RI values were regarded as independent numerical values (in the extreme right of the table).

(ii) The net depicted the statistical aspect of the prediction by RI. On the contrary prediction is 20% to 60% when ADC values or Choline-Creatine ratio were considered (Table 2).

Table 8. Prediction shown by ANN

1.4012	140	3	15	7	6.8	3.55	G Matter		
1.4113	120	3	15	7	6.8	3.55	G Matter		
1.4123	100	3	14	7	6.8	3.52	Wmtter		
1.4144	85	3	14	7	6.8	3.52	Wmtter		
1.4169	70	3	13	7	6.8	3.52	Wmtter		
1.4251	60	3	13	7	6.8	3.52	Wmtter		
1.4288	51	2.8	4.8	3	6	3.51	Giloma		
1.4291	51	2.8	4.8	3	6	3.51		predict	Giloma
1.4311	51	2.8	4.8	3	6	3.51		predict	Giloma
1.4315	51	2.8	4.8	3	6	3.51		predict	Giloma
1.4321	51	2.8	4.8	3	6	3.51		predict	Giloma
1.4435	45	1	3	3.2	10	1		predict	Glioblastoma
1.4439	41	1	3	3.2	10	1		predict	Glioblastoma
1.4446	38	1	3	3.2	11	1		predict	Glioblastoma
1.4551	36	1	3	3.2	11	1		predict	Glioblastoma
1.4624	34	10	3	1.8	1	1.8		predict	Lymphoma
1.4676	32	10	3	1.8	1	1.8		predict	Lymphoma
1.4782	31	10	3	1.8	1	1.8		predict	Lymphoma
1.4799	29	10	3	1.8	1	1.8		predict	Lymphoma
1.4834	28	12	2	1.8	1	1.8		predict	METS
1.4911	28	12	2	1.8	1	1.8		predict	METS

(iii) **Cross Validation:** The aim of cross-validation is to justify the network's capability to predict new data that were not used during training and in order to detect issues like over fitting and to provide an insight on how the net will work to an independent dataset (Table 8).

Table 9.Showing the results of 10 fold cross validation method for the data (Ref.5)

Sample Number	No. of incorrect prediction (out of 24)	Classification rate (in %)	Sensitivity (in %)	Specificity (in %)
1	4	83.33	75	85
2	3	87.5	75	90
3	6	75	71.43	76.47
4	4	83.33	75	87.5
5	1	95.83	100	95
6	3	87.5	100	85
7	3	87.5	83.33	88.89
8	2	91.67	100	89.47
9	4	83.33	80	84.21
10	3	87.5	83.33	88.89

In this Table 9 number of classification rate, incorrect prediction, sensitivity and specificity have been depicted using 10 fold cross validation method. The classification rate observed using 10 folds was quite high and very few errors have been observed in the prediction of test samples (10,11). Corresponding sensitivity and specificity have also been shown in the table. The error is very little between 0.15 to 0.2 units. Thus the dataset has been trained such that the prediction error reaches a minimum value and then testing has been conducted using this trained model.

B. **S V M:** On a data set of 19 patients (Table7) the trained SVM data set was run. For logical reasons these 19 data sets were kept out of the usual training set for prediction purpose. Original biopsy proven diagnoses were tallied with the predicted dataset after running the code. The SVM classified each of the 19 data sets accurately. No erroneous classification was encountered. Several tests produced accurate results constantly with 0% false classifications.

Thus, this Support Vector Machine enabled code correctly discriminates the different types of malignant and benign brain lesions effectively. It clearly characterizes normal gray/white matters, CSF and pathological lesions as well. Results are recorded in Table 10.

Table 10: SVM Prediction Results (100% Accuracy for given data set)

ORIGINAL DIAGNOSIS	RESULT BY SVM	PREDICTION ACCURACY
CSF	CSF	ACCURACY 100%
CSF	CSF	ACCURACY 100%
MS	MS	ACCURACY 100%
MS	MS	ACCURACY 100%
G MATTR	G MATTR	ACCURACY 100%
G MATTR	G MATTR	ACCURACY 100%
W MATTER	W MATTER	ACCURACY 100%
W MATTER	W MATTER	ACCURACY 100%
EDEMA	EDEMA	ACCURACY 100%
EDEMA	EDEMA	ACCURACY 100%
GLIOMA	GLIOMA	ACCURACY 100%
GLIOMA	GLIOMA	ACCURACY 100%
Gblastma	Gblastma	ACCURACY 100%
Gblastma	Gblastma	ACCURACY 100%
METS	METS	ACCURACY 100%
METS	METS	ACCURACY 100%
METS	METS	ACCURACY 100%

CONCLUSION:

It can be summarized that application of Supervised Machine Learning through ANN and SVM (16) is extremely reliable method for accurate diagnosis particularly where imaging techniques and MRS graphs are confusing and misleading. Thus a stereotaxic biopsies which have potential risks to the patient can be avoided (17).

From the literature it was learned that ANN often has the emphasis on local minima rather than the global minima, indicating that they basically miss the “big thing” frequently and thus SVM has edge over ANN. But in this study using Neural Tool 7.5, success rate is extremely high and at par with the performance of SVM. ANN learns slowly but predicts rapidly and has very lightweight models on the other hand SVMs learn rapidly and predict slowly.

REFERENCES:

1.Biswas TK, Gupta A.Retrieval of true color of the internal organ of CT images and attempt to tissue characterization by refractive index : Initial experience. Indian Journal of Radiology and Imaging 2002;12:169-178.

2 Mehryar Mohri, Afshin Rostamizadeh, Ameet Talwalkar (2012) Foundations of Machine Learning, The MIT Press ISBN 9780262018258.

3. Stuart J. Russell, Peter Norvig (2010) Artificial Intelligence: A Modern Approach, Third Edition, Prentice Hall ISBN 9780136042594.

4. Biswas TK, Luu T, In vivo MR Measurement of Refractive Index, Relative Water Content and T2 Relaxation time of Various Brain lesions With Clinical Application to Discriminate Brain Lesions. The Internet Journal of Radiology 2009;13(1).

5. T K Biswas, R Bandopadhyay, A Dutta, Validating The Discriminating Efficacy Of MR T2 Relaxation Value Of Different Brain Lesions And Comparison With Other Differentiating Factors: Use Of Artificial Neural Network And Principal Component Analysis. The Internet Journal of Radiology. 2017 Volume 20 Number 1. ISPUB DOI: 10.5580/IJRA.52614

6. T K Biswas, S R Choudhury, A Ganguly, R Bandopadhyay, A Dutta, Refractive Index As Surrogate Biological Marker Of Tumefactive And Other Form Of Multiple Sclerosis And Its Superiority Over Other Methods, Internet Journal of Radiology, https://print.ispub.com/api/0/ispub-article/46167.

7.Kono K, Inoue Y, Nakayama K, et al. The role of diffusion-weighted imaging in patients with brain tumors. AJNR Am J Neuroradiol 2001; 22: 1081–1088.

8.Pauleit Dirk, Langen Karl-Josef,Floeth Frank, Markus J Riemenschneider, Reifenberger Guido, Shah N. Jon ,Müller Hans-Wilhelm Can the apparent diffusion coefficient be used as a noninvasive parameter to distinguish tumor tissue from peritumoral tissue in cerebral gliomas?.J Magn Reson Imaging. 2004 Nov;20(5):758-64.

9. G. James Variance and Bias for General Loss Functions, Machine Learning 2003; 51, 115-135.(http://www-bcf.usc.edu/~gareth/research/bv.pdf

10. Haykin S., Neural Networks: A Comprehensive Foundation, 2nd edition, Pearson Educ. Asia, Hong Kong, 2001.

11. Bishop C.M., Neural Networks for Pattern Recognition, Clarendon Press, Oxford, 1995.

12.Text Book of Support Vector Machines (SVM), http://www.statsoft.com/Textbook/Support-Vector-Machines

13. Mingxia Liu,Daoqiang Zhang ,Songcan Chen.Hui Xue,Joint Binary Classifier Learning for ECOC-Based Multi-Class Classification IEEE Transactions on Pattern (2015),2335 – 2341

14. PanosM.Pardalos,Hyperplane Arrangement in Optimization Srpringer, Boston,MA,ISBN 978-0-387-74758-3.

15. Support Vector Machines for Binary Classification https://in.mathworks.com/help/stats/support-vector-machines-for-binary- classification.html

*16. Samuel, Arthur, Some Studies in Machine Learning Using the Game of Checkers. IBM Journal of Research and Development,1959;Vol **3**(3): 210–229*

17. Wells, S Lillian, Stereotaxic Brain Biopsy, https://neurosurgery.ufl.edu/residency/about-us/clinical- specialties/stereotactic-brain-biopsy/

18. Michal Antkowiak, Artificial Neural Networks vs. Support Vector Machines for Skin Diseases Recognition, Master's Thesis in Computing Science, Department of Computing Science, Ume°a University 2006 https://www.iitk.ac.in/eeold/archive/courses/2013/intel-info/d4pdf2.pdf