SOCIETY FOR SCIENCE AND EDUCATION UNITED KINGDOM

JBEMi JOURNAL OF BIOMEDICAL ENGINEERING AND MEDICAL IMAGING

ISSN: 2055-1266
VOLUME 5 ISSUE 3

Diagnosis of Brain Lesions, Glioma, Multiple-Sclerosis and Metastases from MRI: An efficient classifier-aided method using Refractive Index as a surrogate Biological Marker

Sparsh Jain, [1,2]Tapan K Biswas, [1]Rajib Bandyopadhyay
[1]Department of Instrumentation and Electronics Engineering, Jadavpur University, Kolkata, India;
[2](Former) Harvard University, Brigham and Women's Hospital, Boston MA, United States of America;
sparshju@gmail.com

ABSTRACT

We introduce a highly accurate method of diagnosing the various pathological conditions that might exist in a subject's brain, like edema, multiple sclerosis (Tumefactive, Relapsing-remitting, secondary-progressive, primary-progressive and progressive-relapsing), glioma, glioblastoma and metastases. These show up on conventional MRI scans, but it is often difficult to identify the exact type of the pathology from the grayscale image. We employ the use of Support Vector Machines (SVM) to work on the MR Spectroscopy [6, 12] data and correctly identify the condition-especially in seemingly vague cases where radiologists cannot rule out high uncertainty in their conclusion. The SVM trains on data sets collected for different patients and optimizes its hyperplanes based on eight input variables – T2, CHO, ADC, CR, CHO/NAA, CR/NAA, LIP/LAC, MI, CH/CR, T2 periphery [6] and Refractive index. Refractive index is an additional parameter which we include to get better boundary lines and accuracy, as shown in our prior works [10]. We test this SVM on a set of 19 patients' data and achieve 100% accuracy in predictions. The training and testing is carried out in MATLAB.

Keywords: Magnetic Resonance Imaging, MR Spectroscopy, Refractive Index, Support Vector Machine (SVM), Brain Lesions, Cancer.

1 Introduction

Only structural changes of MR images of aneurysms, Glioma or other brain tumours including metastasis or secondary deposits of cancer tissues cannot be sufficient for accurate diagnosis [1]. Data collected from metabolites of MR spectroscopy like NAA, Choline, Creatine, lipid or lactate or physical data like Refractive Indices (RI), T2 magnetic relaxation values, and Apparent Diffusion Coefficient (ADC) values are also important for correct diagnosis of the disease. Figures 1 to 4 show how a few of these diseases appear on the scans [Fig 1-4].

If the supporting data are available, live prediction of diseases or of the tissue can be performed with 90 to 95% accuracy using Support Vector Machines (SVM). Clustering of diseases and tissues can also be done using SVMs. It is worth mentioning that machine learning tools like clustering, SVM and neural

DOI: 10.14738/jbemi.53.4700
Publication Date: 19th June 2018
URL: http://dx.doi.org/10.14738/jbemi. 53.4700

networks (Principal Component Analysis) have previously been used for similar purposes on MRI data [5, 10, 13-15].

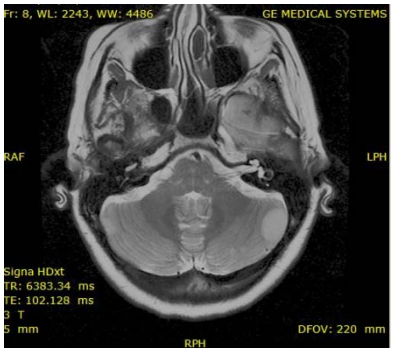

Figure 1: Metastasis MRI

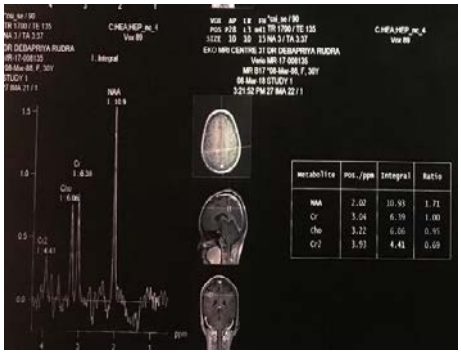

Figure 2: MR Spectroscopy (MRS) Normal

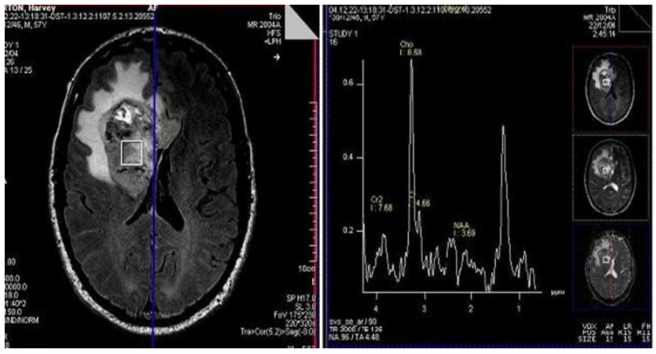

Figure 3: Grade-3 Glioblastoma with abnormal MRS

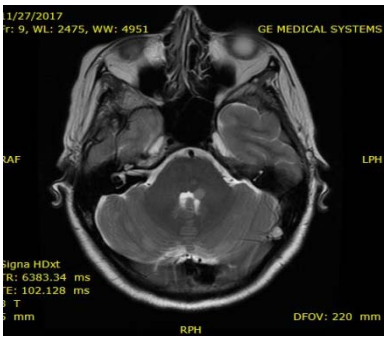

Figure 4: Multiple Sclerosis (MS) MRI

2 Background

SVM employs a nonlinear classification method [2]. It is implemented to assess and make virtual pathological predictions using the data obtained from MRI, various metabolic components and their ratio of MR Spectroscopy, ADC, RI and T2 values (Table 1).

An SVM with extraordinary data processing uniqueness, nonlinearity and learning with generalization capability is used to characterize the disease. Thus there are 10 independent numeric variables. SVM and Neural Network (NN) both belong to supervised learning methods, but their working procedure is different.

We have used Error correcting output codes (ECOC) mode to reduce the errors in classification. ECOC is an error-correcting output codes classifier used to get multiclass learning by diminishing multiple binary classifiers for instance SVMs [3].

The main factor responsible for the performance of ECOC methods is the self-determination of binary classifiers, otherwise the ECOC method will be unsuccessful. It is effective in multiple classes and requires a coding design. This design determines the classes. Binary learners undergo training and a decoding system determines the prediction of the binary classes.

a) The coding design is one-versus-one.

b) Decoding procedure would utilize loss *g*.

c) SVMs will be the learners.

SVM utilizes a function or hyperplane [4] among different classes giving a maximum margin parameter. This hyperplane tries to divide the classes so that each groups remains on either side of the plane and by a particular boundary. Our SVM handles 10 variables. It has 8 output results or classes of different types of tissue (such as CSF, gray matter, white matter) and diseases like low and high grade glioma and metastasis for targeted prediction.

Improved classification accuracy can be achieved by ECOC models compared to other multiclass model.

3 Method

The data has been used from the authors' prior research [5] (Table1). We train the SVM using data from 116 patients. The SVM is then tested on data from 19 patients.

The method consists of two parts:

a) Training the SVM using available data

b) Testing the trained SVM to classify unknown data.

3.1 Source Code and Platform

The SVM training and classification work has been implemented on the 64-bit MATLAB R2017a environment on Windows 10 Home platform. Access to the source code and complete training data can be requested by contacting the corresponding author.

3.2 Training the Support Vector Machine

3.2.1 3.2.1 Identification

First we identify the label and data. Label includes the names of tissue or diseases like MS, Glioma etc. Data corresponds to all the numerical values associated with each label. SVM essentially matches the data sets with the correct labels. The first column of Table 1 has all the labels. The remaining 11 columns contain the data. So for example, row no. 5 has the label CSF and that particular CSF scan has 11 values corresponding to T2, ADC, CR, CHO etc [6].

Keys for tables 1-4:

'CSF': Cerebrospinal fluid

'MS': Multiple sclerosis

'gmatter': Gray Matter

'w matter': White Matter

'gblastma': Glioblastoma

'mets': Metastases

Table 1: Preview of the entire data set. 41 out of 135 sets are shown below.

TISSUE	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	394	1710	322	1809	0.423	0.368	1690	1059	1.14	394	1.3333
CSF	345	2021	402	2060	0.572	0.448	1744	1145	1.15	345	1.3333
CSF	344	2022	403	2061	0.573	0.451	1744	1145	1.15	344	1.3333
CSF	343	2023	404	2062	0.574	0.452	1745	1146	1.15	343	1.3333
CSF	342	2024	405	2068	0.577	0.453	1746	1147	1.15	342	1.3333
CSF	341	2123	411	2063	0.578	0.453	1747	1148	1.15	341	1.3333
ms	340	11750	145	8320	0.779	0.557	4160	2912	1.4	340	1.3334
ms	339	11750	1460	8319	0.778	0.541	4423	3223	1.4	339	1.3335
ms	338	11749	1459	8314	0.776	0.538	4423	3221	1.4	338	1.3336
ms	337	11746	1445	8311	0.774	0.536	4421	3220	1.4	337	1.3421
ms	336	11745	1444	8310	0.773	0.534	4422	3219	1.4	336	1.3439
ms	335	11745	1443	8309	0.772	0.532	4420	3216	1.4	335	1.3498
ms	334	11743	1443	8308	0.771	0.531	4419	3214	1.4	334	1.3499
ms	304	5947	120	5400	0.873	0.7396	6766	4294	1.1	245	1.3519
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
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	93	1067	89	2301	0.444	0.775	1167	324	0.466	169	1.4262
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	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	99	1365	177	2254	0.341	0.712	917	300	0.343	170	1.4339
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	128	1284	131	2589	0.541	0.781	1767	322	0.76651	198	1.4723
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

3.2.2 Training

We use the Fitcecoc command in Matlab to train an SVM for the data and label we already have (Table 2). SVM uses Supervised Learning and classifies data. The basic command for SVM training is FitSVM or svmtrain but that can be used only for binary classification i.e. when there are only two classes/labels [7]. In this case we have 8 different labels, so we use fitcecoc, which can accommodate multiple classes. SVM and Neural networks both work on supervised learning but their mechanisms differ.

So the output variable generated is the Support Vector Machine trained using the data.

Table 2: Preview of the Training data set. 16 out of 116 sets are shown below.

TISSUE	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	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	340	1.3334
ms	336	11745	1444	8310	0.773	0.534	4422	3219	1.4	336	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	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
edema	160	1231	132	2216	0.443	0.776	1123	325	0.467	246	1.3741
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
Gblastma	108	2457	154	2115	0.332	0.676	900	311	0.311	195	1.4512
Gblastma	128	1284	131	2589	0.541	0.781	1767	322	0.76651	198	1.4723
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

3.3 Testing the SVM to predict the classes of unknown data sets

Using a few data sets (which were not used to train the SVM) we test the SVM. The data (Table 3) to the SVM is fed and it predicts the classes on its own.

Table 3: Test data set. The trained SVM predicts the classes for 19 unknown data sets.

TISSUE	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	395	1711	322	1800	0.422	0.367	1701	1056	1.14	395	1.3333
CSF	394	1710	322	1809	0.423	0.368	1690	1059	1.14	394	1.3333
ms	337	11746	1445	8311	0.774	0.536	4421	3220	1.4	337	1.3421
ms	335	11745	1443	8309	0.772	0.532	4420	3216	1.4	335	1.3498
ms	334	11743	1443	8308	0.771	0.531	4419	3214	1.4	334	1.3499
ms	304	5947	120	5400	0.873	0.7396	6766	4294	1.1	245	1.3519

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	1067	89	2301	0.444	0.775	1167	324	0.466	169	1.4262
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
METS	129	1298	130	2567	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

4 Results

We ran our trained SVM on a data set of 19 patients. These 19 sets were excluded from the training set for obvious reasons. We had the original diagnoses prior to running the code. The SVM successfully classified each of the 19 data sets accurately. No misclassification was encountered. We retrained the SVM by taking different data sets out of the training set and using them for test purposes, owing to lack of new data. Multiple tests yielded consistently accurate results i.e. the SVM perfectly classified the data with 0% uncertainty or false classifications. Thus, this Support Vector Machine enabled code correctly diagnoses the seven different types of brain diseases reliably. It also clearly differentiates between normal CSF tissue and lesions/pathological conditions.

Results are recorded in Table 4.

Table 4: Final results comparing the SVM prediction to the actual diagnoses.

ORIGINAL PATIENT DIAGNOSIS	RESULT PREDICTED BY OUR SVM	ACCURACY OF PREDICTION
'CSF'	'CSF'	Accurate Classification
'CSF'	'CSF'	Accurate Classification
'ms'	'ms'	Accurate Classification
'ms'	'ms'	Accurate Classification
'gmatter'	'gmatter'	Accurate Classification
'gmatter'	'gmatter'	Accurate Classification
'w matter'	'w matter'	Accurate Classification
'w matter'	'w matter'	Accurate Classification
'edema'	'edema'	Accurate Classification
'edema'	'edema'	Accurate Classification
'GLIOMA'	'GLIOMA'	Accurate Classification
'GLIOMA'	'GLIOMA'	Accurate Classification
'Gblastma'	'Gblastma'	Accurate Classification
'Gblastma'	'Gblastma'	Accurate Classification
'Gblastma'	'Gblastma'	Accurate Classification
'METS'	'METS'	Accurate Classification
'METS'	'METS'	Accurate Classification
'METS'	'METS'	Accurate Classification
'METS'	'METS'	Accurate Classification

5 Conclusion

It can be concluded that the use of Support Vector Machines and other tools of Machine Learning [8] can be employed to achieve extremely reliable and accurate diagnostic results in cases where conventional imaging techniques offer less clarity, and where biopsies [9] carry potential risks to the

patient. An added advantage of using SVM for this purpose is that it guarantees reproducibility and repeatability i.e. the accuracy levels are guaranteed to remain fairly constant over multiple runs and long periods of time. Along with the custom-developed colour palette [10] using refractive index, it also provides a much sharper contrast image of the brain. We here note that the accuracy is limited only by the size of the data set we use for training the SVM initially. Owing to our present setup, we are unable to record new data on a rolling basis. A further step towards making it more reliable and robust would include obtaining new data in large amounts, and incorporating these data points into the training set. Different classification algorithms can further be integrated with this to achieve even better and conclusive results in all cases. It is safe to declare that Support Vector Machines will find their use in day-to-day radiological and imaging methods for common use due to their superior accuracy, fail-proof safety fallbacks and evolving classification algorithms. This in turn will slowly reduce the need to perform invasive biopsies and multiple tests, as well as make diagnoses less dependent on the human radiologist/ practitioner who might be prone to making errors [11].

REFERENCES

[1] Schwab, K.E., Gailloud P, Wyse G, Tamargo R.J., *Limitations of magnetic resonance imaging and magnetic resonance angiography in the diagnosis of intracranial aneurysms*. *Neurosurgery*, Volume 63, Issue 1, 1 July 2008

[2] Support Vector Machines - http://www.statsoft.com/Textbook/Support-Vector-Machines

[3] Mingxia Liu, Daoqiang Zhang, *Joint Binary Classifier Learning for ECOC-Based Multi-Class Classification*. IEEE Transactions on Pattern Analysis and Machine Intelligence (Volume: 38, Issue: 11, Nov. 1 2016)

[4] Pardalos P.M. (2008) *Hyperplane Arrangements in Optimization*. In: Floudas C., Pardalos P. (eds) Encyclopedia of Optimization. Springer, Boston, MA

[5] Tapan K Biswas, R Bandyopadhyay, 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 (Volume 20, No. 1)

[6] Radiopedia Introduction to MR Spectroscopy, https://radiopaedia.org/articles/mr-spectroscopy-1.

[7] Mathworks Guide, Support Vector Machines for Binary Classification.

[8] Machine Learning, Wikipedia, https://en.wikipedia.org/wiki/Machine_learning.

[9] Nor M., Noor Rahman, J. Adnan, *Intracranial Bleed Post Stereotactic Biopsy: Lessons Learned*, The Internet Journal of Neurosurgery (Volume 8, Number 1)

[10] Tapan K Biswas, S R Choudhury, A Ganguly, R Bandyopadhyay, A Dutta, *Refractive Index As Surrogate Biological Marker Of Tumefactive And Other Form Of Multiple Sclerosis And Its Superiority Over Other Methods*. The Internet Journal of Radiology (Volume 19, Number 1).

[11] Blake A. Johnson, *Avoiding diagnostic pitfalls in neuroimaging.* Applied Radiology-The Journal of Practical Medical Imaging and Management 2016;45(3):24-29. March 02, 2016

[12] Wikipedia, Nuclear Magnetic Resonance Spectroscopy, https://en.wikipedia.org/wiki/Nuclear_magnetic_resonance_spectroscopy

[13] Kasai, M., Yasuda, Y., Mizoguchi, H., Soga, K., Kaneko, K., & Takemura, H. (2017). *In vivo tumor wavelength band selection using Hierarchical clustering and PCA with NIR-Hyperspectral Data*. Journal of Biomedical Engineering and Medical Imaging, *4*(1), 01.

[14] Limam, O. (2016). *MRI Segmentation based on Multiobjective Fuzzy Clustering*. Journal of Biomedical Engineering and Medical Imaging*, 3*(2), 07.

[15] Abd El kader, I., Zhang, S., & Xu, G. (2017). *Improved Fuzzy C-Means Algorithm for Brain Tumor Identification Analysis Using Magnetic Resonance Brain Images*. Journal of Biomedical Engineering and Medical Imaging, 4(3), 15.

