



Multiclass Tumour classification by using SVM classifiers

P.K.Srimani¹ and Shanthi Mahesh²

¹Dept. of Computer Science & Maths,

Bangalore University, Director R&D, BU. Bangalore-560078, Karnataka, India.

²Dept. of Information Science & Engineering,

Atria Institute of Technology, Bangalore-560024, Karnataka, India.

Abstract: Brain cancer is a disease in which cells grow in an uncontrolled fashion in the brain. Therefore, a proper diagnosis of the disease is absolutely required so that effective treatment could be provided to the patients. This paper presents brain tumour classification of five classes using SVM classifier. The classes considered are Medulloblastoma, Malignant glioma, Normal cerebellum and PNET from the data set Brain_tumor1.mat), which includes 90 sample, each sample has 5920 genes. From the experiment conduct in Matlab using SVM classifier, the result was found to be Normal cerebellum has sensitivity of 1, specificity of 0.5000 and prevalence of 95.6%. This chapter presents brain tumour classification of five classes using SVM classifier. The classes considered are Medulloblastoma, Malignant glioma, Normal cerebellum and PNET from the data set Brain_tumor1.mat), which includes 90 sample, each sample has 5920 genes. From the experiment conduct in Matlab using SVM classifier, the result was found to be Normal cerebellum has sensitivity of 1, specificity of 0.5000 and prevalence of 95.6%.

Key Words: Brain Tumour, SVM classifier, Sensitivity, Specificity, Prevalence.

I. Introduction

Tumour is basically defined as the abnormal growth of the tissues. Brain tumour is an abnormal mass of tissue in which cells grow and multiply uncontrollably. Brain tumours can be primary or metastatic and either benign or malignant. A metastatic brain tumour is a cancer that can spread from anywhere in the body to the brain.

The knowledge of volume of a tumour plays an important role in the treatment of malignant tumours. Manual segmentation of brain tumours from Magnetic Resonance images is a challenging and time consuming task. This paper presents a novel technique for the detection of tumour in brain using segmentation and histogram thresholding. The proposed method can be successfully applied to detect the contour of the tumour and its geometrical dimensions. This technique is proved to be a handy tool for physicians engaged in this field.

Brain cancer is one among the most deadly and intractable diseases. Tumours may be embedded in regions of the brain that are critical to orchestrating the body's vital functions, while they shed cells to invade other parts of the brain, forming more tumours too small to detect using conventional imaging techniques. Brain cancer's location and ability to spread quickly makes treatment with surgery or radiation like fighting an enemy hiding out among minefields and caves. Brain cancer is a disease in which cells grow uncontrollably in the brain. Brain tumours are of two main types: (i) Benign tumours (ii) Malignant tumours. Benign tumours are incapable of spreading beyond the brain itself. Benign tumours in the brain usually do not need to be treated and their growth is self limited. Sometimes they cause problems because of their location, and surgery or radiation can be helpful. Malignant tumours are typically called brain cancer. These tumours can spread outside the brain. Malignant tumours of the brain are most harmful which may remain untreated and an aggressive approach is almost always warranted. Detection of Brain tumour is a serious issue in medical science. Brain tumour is one of the major causes for the increase in mortality among children and adults. Imaging plays a central role in the diagnosis and treatment planning of brain tumour. Imaging of the tumours can be done by CT scan, Ultrasound and MRI etc. The MR imaging method is the best due to its higher resolution. But there are many problems in the detection of brain tumour in MR imaging as well. An important step in most medical imaging analysis systems is to extract the boundary of an area we are interested in. Many of the methods are there for the MRI segmentation. Till today histogram thresholding is used for pre-processing only in many of the segmentation methods. This paper shows that it can be used as a powerful tool for segmentation also. The image captured from an affected tumours brain shows the place of the infected portion of the brain. The image does not give the information about the numerical parameters such as area and volume of the infected portion of the brain. However, after the preprocessing of the image, the image segmentation is done first. The segmented image shows the unhealthy portion clearly. From this image the infected portion (tumour) is selected by cropping the segmented image.

From this cropped image, statistical analysis viz., mean, standard deviation, correlation and the area or dimension of the tumour is calculated.

II. Related work

In this section, a brief review of the literature is presented. Quite a good amount of literature pertaining to the application of SVM classification different area is available. Authors [1] have proposed an approach that uses mirror point pairs and a multiple classifier system to reduce the classification time of a SVM. A greedy strategy is then used for fine selection of member classifiers. Authors[2] have proposed a parallelization of support vector machine learning for shared memory systems. The learning algorithm relies on a decomposition scheme, which in turn uses a special variable projection method, for solving the quadratic program associated with SVM learning. Authors[3], have presented Pattern recognition approaches, such as the SVM, have been successfully used to classify groups of individual based on their patterns of brain activity. This approach is an application of one-class SVM to investigate if patterns of fMRI response to sad facial expression in depression patients would be classified as outliers in relation to pattern of health control subjects. Authors[4] presented a SVM classifier based on Rough Set(RS) is researched in order to enhance the predicting performance. Authors[5], have used Diffusion tensor imaging (DTI) in conjunction with machine learning algorithms in order to automate the classification of healthy older subjects and subjects with mild cognitive impairment(MCI). This method shows the potential of a DTI and SVM pipeline for fast, objective classification of healthy older and MCI subjects. Authors[6], have made a study to use SVM in classifying e-mail as spam or nonspam by comparing it to three other classification algorithms and shown that SVM's had significantly less training time. Authors[7] have presents a novel learning method, SVM, is applied on different data which have two multiclass and comparative results are shown using different kernel function. Authors[8], have proposed a primary method to overcome SVM drawback with respect to classification speed which is due to the number of support vectors being large. Authors[9] , have presented and implement SVM using R tool. Authors[10], have introduced an alternative implementation technique for SVM to address the classification problem with small-size training sample set. It has been empirically proven that the effectiveness of the introduced implementation technique which has been evaluated by benchmark dataset. Authors[11], have proposed a convergence of a generalized version of modified SMO for SVM classifier design is proved, the results are also extended to modified SMO algorithms for solving v-SVM classifier problem. Authors[12] have used machine learning approaches to analyze multi-dimensional experimental data and integrate results from many studies, yet they have not been applied to specifically tackle the task of predicting and classifying self renewal and pluripotency gene memberships.

III. Dataset Description

BrainTumour Data (Brain_Tumor1.mat)

The dataset comes from a study of 5 human brain tumor types and includes 90 samples. Each sample has 5920 genes:

- a. **Medulloblastoma:** Medulloblastoma is a highly malignant primary brain tumor that originates in the cerebellum or posterior fossa. Previously, medulloblastomas were through to represent a subset of primitive neuroectodermal tumor (PNET) of the posterior fossa. However, gene expression profiling has shows that medulloblastomas have distinct molecular profile and are distinct from other PNET tumors.
- b. **Malignant glioma: A glioma is a type of malignant brain tumor.** A malignant tumor is a mass of abnormal cells that is cancerous.
Tumors can develop in any part of the brain or its nerves and covering tissues. The two major types of brain tumor are primary and secondary. Primary brain tumors start in the brain. Secondary brain tumors start in another part of the body, then spread to the brain. A glioma is a primary brain tumor, accounting for 45% of cancers that begin in brain cells.
The three main types of glioma include: astrocytoma, ependymoma, and oligodendroglioma. Each of these types can be assigned a grade, either low grade or high grade, with high grade being more malignant and aggressive.
Astrocytomas are named for the cells where they originate, the astrocytes. These tumors can either show clear borders between normal brain tissue and the tumor (called focal) or no clear border (called diffuse). Focal astrocytomas are most common in children and are not often found in adults.
Ependymomas begin in cells called ependymal cells that are found lining certain areas of the brain and spinal cord. These cells help repair damaged nerve tissue. They usually occur in children and young adults.
Oligodendrogliomas form in oligodendrocyte cells, which produce a fatty substance called myelin that protects the nerve. More common in adults, these tumors may move to other parts of the brain or spinal cord.
- c. **AT/RT (atypical teratoid/rhabdoidtumours):**

Atypical teratoid/rhomboid tumor (AT/RT) is a rare tumor usually diagnosed in childhood. Although usually a brain tumor, AT/RT can occur anywhere in the central nervous system (CNS) including the spinal cord. About 60% will be in the posterior cranial fossa (particularly the cerebellum). Figure 1 shows MRI of AT/RT and figure 2 shows the cerebellum with surrounding skull and spinal fluid occupies the bottom 1/3 of the axis of MRI image.. One review estimated 52% posterior fossa, 39% sPNET (supratentorial primitive neuroectodermal tumors), 5% pineal, 2% spinal, and 2% multi-focal.

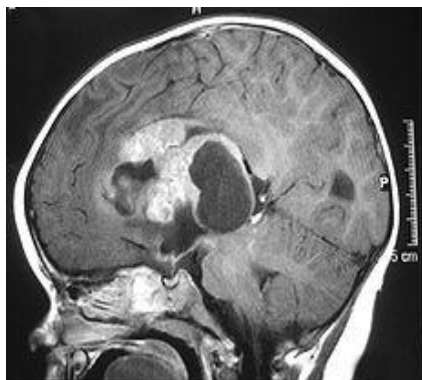


Figure 1: MRI of AT/RT

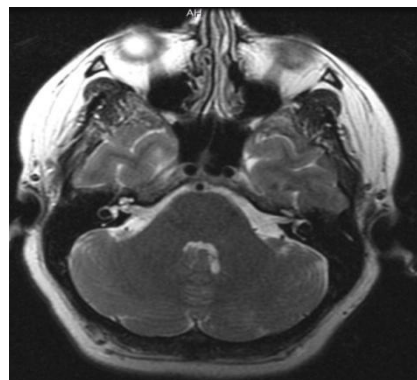


Figure 2: The cerebellum with surrounding skull & spinal fluid occupies the bottom 1/3 of this axis MRI

d. Normal cerebellum

The cerebellum is part of the brain. It lies under the cerebrum, towards the back, behind the brainstem and above the brainstem. The cerebellum is largely involved in "coordination". Persons whose cerebellum doesn't work well are generally clumsy and unsteady. They may look like they are drunk even when they are not.

The main clinical features of cerebellar disorders include incoordination, imbalance, and troubles with stabilizing eye movements. There are two distinguishable cerebellar syndromes -- midline and hemispheric. Midline cerebellar syndromes are characterized by imbalance. Persons are unsteady, they are unable to stand in Romberg with eyes open or closed, and are unable to well perform tandem gait. Severe midline disturbance causes "trunkal ataxia" a syndrome where a person is unable to sit on their bed without steadying themselves. Some persons have "titubation" or a bobbing motion of the head or trunk. Midline cerebellar disturbances also often affect eye movements. There may be nystagmus, ocular dysmetria and poor pursuit.

Hemispheric cerebellar syndromes are characterized by incoordination of the limbs. There may be decomposition of movement, dysmetria, and rebound. Dysidiadochokinesis is the irregular performance of rapid alternating movements. Intention tremors may be present on an attempt to touch an object. A kinetic tremor may be present in motion. The finger-to-nose and heel-to-knee tests are classic tests of hemispheric cerebellar dysfunction. While reflexes may be depressed initially with hemispheric cerebellar syndromes, this cannot be counted on. Speech may be dysarthric, scanning, or have irregular emphasis on syllables.

e. PNET(primitive neuroectodermaltumours):

PNET(pronounced pee-net) stands for a group of tumours known as Primitive Neuro Ectodermal Tumours. PNETS develop from cells that are left over from the earliest stages of a body's development in the womb.

Normally these cells are harmless. But occasionally they turn into a cancer. These cancers are more common in children than adults.

Doctors use the term PNET to classify the tumour. They are divided into two main groups:

- a. PNETs of the brain and central nervous system
- b. Peripheral PNET (outside the brain and nervous system)

PNETs of the brain or spinal cord

Primitive neuroectodermal tumours that occur in the brain and spinal cord include:

- Medulloblastoma (develops in the back part of the brain – the hindbrain)
- Pineoblastoma (develops in the pineal region of the brain)
- Non pineal supratentorial

IV. Methodology

Support vector machines (SVM) is a class of linear classification algorithms which tries to maximize the margin of confidence of the classification on training data set and has been considered as well-suited to the noise and

high-dimensionality of microarray data. In its simplest form, the decision boundary of SVM is a hyperplane that separates boundary training data of different classes with maximum confidence. Unlike LDA, if no separating hyperplane exists, SVMs will map the samples to a higher-dimensional space where the data become linearly separable.

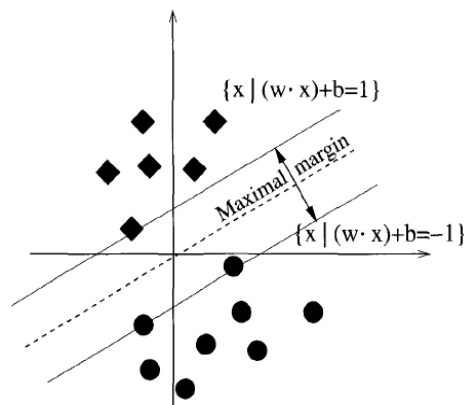


Figure 3: The optimal hyperplane for the separation of balls and diamonds has the maximal margin between the two classes.

SVM classifiers are based on the class of hyperplanes $(w \cdot x) + b = Q, w \in \mathbb{R}^N, b \in \mathbb{R}$, corresponding to decision functions $f(x) = \text{sign}((w \cdot x) + b)$. This is a convex, quadratic programming problem which is approached using Lagrangian. Figure (3) shows that the optimal hyperplane is defined as that with the maximal margin of separation between the two classes.

SVMs map the data into some other dot-product space (called the feature space) F via a nonlinear map $\Phi: \mathbb{R}^N \rightarrow F$, and try to solve the same optimization problem in F . The mapping is not explicitly needed if some kernel function K is used such that

$$\Phi(i) \cdot \Phi(j) = K(x_i, x_j).$$

The mapping is accomplished by raising the dot-product kernel $x \cdot y$ to a positive integer power. For instance, squaring the kernel yields a convex surface in the input space. Raising the kernel to higher powers yields polynomial surfaces of higher degrees. SVMs are more powerful than linear discriminant methods, since they are able to classify non-linearly separable classes by mapping the samples into higher-dimensional space. Moreover, SVMs are different from other nonlinear classifiers in that they pay special attention to the boundary of separation between the regions corresponding to each class, and this may yield small improvements of the classification prediction rate.

V. Experiments and Results

Experiments were conducted using MATLAB, figure 4a and 4b show variable and Sample profiles of the brain tumour data.

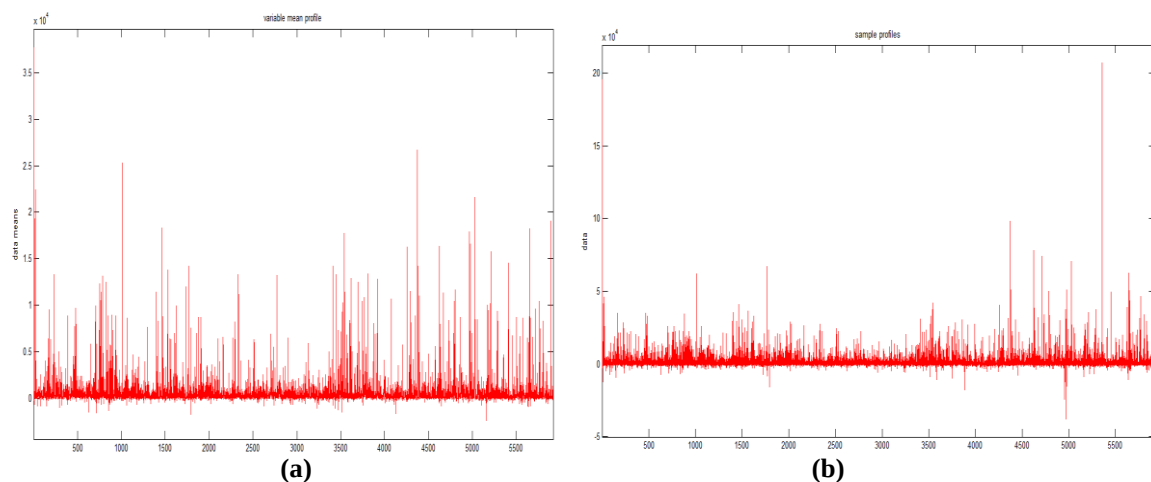


Figure 4: (a) Sample Profile, (b) Variable Profile

Table 1: Classified rate, sensitivity and specificity prediction

Sl.no	Tumor type	Last Error Rate	Classified Rate	Sensitivity	Specificity
1	Medulloblastoma	0.3556	1	0.0667	0.9333
2	Malignant glioma	0.1111	1	1	0
3	AT/RT	0.1111	1	1	0
4	Normal cerebellum	0.0222	1	1	0.5000
5	PNET	0.1111	1	0.9524	0

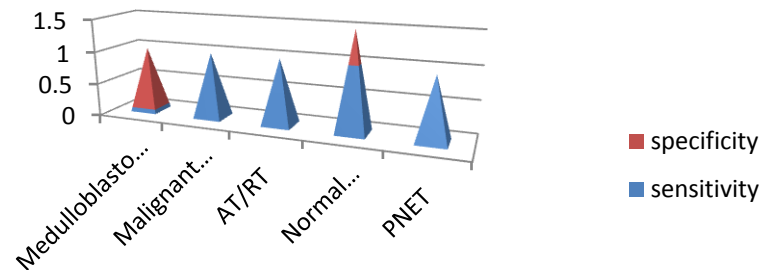


Figure 5: Comparison of specificity and sensitivity of 5 classes of tumor

Figure 5 gives the graphical representation of the results corresponding to the five classes of tumor. Table 2 gives the accuracy (prevalence) prediction while figure 6 gives its graphical representation. The results show that the performance of SVM classifier is excellent.

Table 2: Accuracy Prediction

Sl.no	Tumor type	Positive predictive	NPV	PL	NL	Prevalence
1	Medulloblastoma	0.3333	0.6667	1	1	0.3333
2	Malignant glioma	0.8889	NaN	1	NaN	0.8889
3	AT/RT	0.8889	NaN	1	NaN	0.8889
4	Normal cerebellum	0.9773	1	2	0	0.9556
5	PNET	0.9302	0	0.9524	NaN	0.9333

Form table 1, it is evident that the two tumour types viz., Malignant glioma and AT/RT have the sensitivity equal to 1 and specificity as 0, while Normal cerebellum has sensitivity equal to 1 and specificity equal to 0.5.

Prevalence

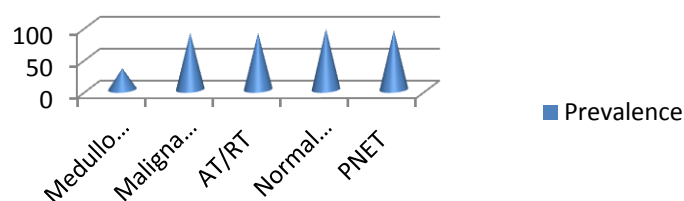


Figure 6: Prevalence comparison

VI. CONCLUSION

Gene expression data can be classified on both genes and samples. As a result, the SVM classification algorithm performs extremely well. From the present investigation it is concluded that SVM classifier is the best classifier

with (Sensitivity=1, specificity=0) when compared to other classification techniques. Normal cerebellum has sensitivity of 1, specificity of 0.5000 and prevalence of 95.6%. The results of this investigation are obtained by implementing the CODE in the MATLAB and they throw light on the qualitative as well as the quantitative aspects of the problem.

VII. References

- [1] Jiun-Hung Chen and Chu-Song Chen, "Reducing SVM classification Time Using Multiple Mirror Classifiers" IEEE Transaction on Systems, Man and Sybernetics, Vol. 34, N0.2, April 2004.
- [2] Tatjana Eitrich and Bruno Lang, "Data Mining with Parallel Support Vector Machine for Classification", ADVIS 2006, LNCS 4243, pp.197-206, 2006 C Springer-Verlag Berlin Heideberg 2006.
- [3] Janaina Maurao, David R Hardoon, Tim Hahh, Andre F Marquand, Steven Williams and Michael Brammer, "Patient classification as an outlier detection problem: an application of the one-class support vector machine", pp.793-804, NeuroImage, Vol.58, 1 Oct 2001.
- [4] Guojun Zhang and Jixiong Chen, "A SVM Classifier Research Based on RS Reducts", Information Management, Innovation Management and Industrial Engineering, 2009 Internatioanl Conference. Dec 26th and 27th 2009. Vol.4
- [5] Laurence O'Dwyer, Arun L. W. Bokde, Michael Ewers, Yetunde O. Faluyi, Colby Tanner, Bernard Mazoyer, Desmond O'Neill, Máiréad Bartley, D. Rónán Collins, Tara Coughlan, David Prvulovic and Harald Hampel, "Using Support Vector Machines with Multiple Indices of Diffusion for Automated Classification of Mild Cognitive Impairment" PLOS ONE, February 23, 2012.
- [6] Harris Drucker, Donghui Wu and Vladimir N.Vapnik, "Support Vector Machines for Spam Categorization", IEEE Transactions on Neural Networks, vol.10, No. 5, September 1999.
- [7] Durgesh KSrivastava and Lekha Bhambhu, "Data Classification Using SVM", 2005-2009 JATIT. WWW.jatit.org
- [8] Sathiya Keerthi, Olovier Chapelle and Dennis DeCoste, "Building Support Machine with Reduced Classifier complexity", Journal of Machine Learning Research 1493-1515, 7 (2006).
- [9] Alexandros Karatzoglou, David Meyer and Kurt Hornik, "Support Vector Machine in R", Journal of Statistical Software, Vol.15, Issue 9, April 2006.
- [10] Mingmin Chi, Rui Feng and Lorenzo Bruzzone. "Classification of hyperspectral remote-sensing data with primal SVM for small-sized training dataset problem" ScienceDirect, Advance in space Research 41(2008) 1793-1799,
- [11] Keerthi S.S and Gilbert E.G, "Convergence of a Generalized SMO algorithm for SVM Classifier Design", Kluwer Academic Publishers, Machine Learning, 46, 351-360, 2002.
- [12] Huilei Xu, Ihor R Lemischka and Avi Maayan. "AVM classifier to predict genes important for self-renewal and pluripotency of mouse embryonic stem cells". BMC System Biology 2010, 4:173.
- [13] Srimani P.K and Shanthi Mahes, "Knowledge Discovery in Image-Segmentation Data set Using Decision Tree Classifiers " International Journal of Current Research, Vol.4, Issue, 09, pp.135-140, september, 2012
- [14] Srimani P.K and Shanthi Mahes, "Knowledge Discovery Process in Image-Segmentation", International Journal of Knowledge Engineering, Vol.3, Issue, 3, Issue 2, pp.188-192, 2012