International Journal of Applied Engineering Research
ISSN 0973-4562 Volume 10, Number 8 (2015) pp. 20451-20459

http://www.ripublication.com

Computer-Aided Classification of Neuroblastoma Using Support Vector Machine

Karamjeet Kaur[1], Sanjay Kumar Singh [1], Anjali Goyal[2]
[1]Department of Computer Science Engineering, Lovely Professional University, Phagwara, Punjab, India
Email: {kaur.karam15; sanjayksingh.012}@gmail.com
[2] Dept. of Computer Applications, GNIMT, Ludhiana
Email: anjali.garg73@gmail.com

Abstract

Neuroblastoma (NB) is the cancer of sympathetic nervous system. NB tumor classification relies on International Neuroblastoma Pathology Classification (INPC) system. In this work important role is played by mitiosis and karyorrhexis index (MKI). Its count helps in determining the histology as favorable or unfavorable. This work mainly relies on Support Vector Machine (SVM) classifier for grading of NB tumor. SVM is applied to the training dataset consisting of different attributes like stage, MYCN amplification, histology, ploidy, MKI, diagnostic category. Based on that, target grading performance is checked. This work when tested with 640 samples showed the best overall accuracy of 99.69%. Much efficient results are obtained which proves this approach effective. This method can be used by doctors for diagnosing the NB tumor earlier and in providing proper treatment to patients.

Keywords: Neuroblastoma, Support Vector Machine, MKI, Stage, International Neuroblastoma Pathology Classification

Introduction

Neuroblastoma is a tumor of early childhood which begins in the nervous system or nerve cells called neuroblasts. It starts from the cells of sympathetic nervous system which is the part of autonomic nervous system. It can be sometimes inherited from parent to the child but in most cases it is not inherited. Around 650 patients are diagnosed with NB each year in United States [1]. Diagnosis of this cancer is carried out by pathologists according to International Neuroblastoma Pathology Classification System developed by Shimada et al. [2]. Its classification is carried out by visual testing of pathological images. According to this classification the tumor is categorized into different groups. This system classifies tumor as undifferentiated

(UD), poorly differentiated (PD) and differentiating (D), depending on stage, histology, MYCN amplification [3], ploidy, diagnostic category [4] and MKI of the patient as described in [2].Histology can be favorable or unfavorable depending on the age and MKI of the patients. Most important step in classifying the histology is determining the mitosis-karyorrhexis index (MKI). It is defined as the number of tumor cells. Mitiosis is the division of a cell into two cells and Karyorrhexis is division of nucleus of dying cell [5]. Based on this index tumor is classified as described in the following tables [6]:

Table 1: Mitiotic figures and karyorrhectic cells/ 5000 neoplastic cells

	Expressed as percent	**Expressed as cell count**
Low MKI	<2%	<100 / 5000 cells
Intermediate MKI	2-4%	100-200 / 5000 cells
High MKI	>4%	>200 / 5000 cells

Table 2: Unfavorable histology

Tumor Subtype	**MKI**	**Patient Age**
Undifferentiated NB	Any MKI	Any
Poorly differentiated NB	High MKI	Any
Poorly differentiated NB	Intermediate MKI	>=1.5 yrs
Poorly differentiated NB	Low MKI	>=1.5 yrs
Differentiating NB	High MKI	Any
Differentiating NB	Intermediate MKI	>=1.5 yrs
Differentiating NB	Not Applicable	>=5 yrs

Table 3: Favorable histology

Tumor Subtype	**MKI**	**Patient Age**
Poorly differentiated NB	Intermediate MKI	<1.5 yrs
Poorly differentiated NB	Low MKI	<1.5 yrs
Differentiating NB	Intermediate MKI	<1.5 yrs
Differentiating NB	Low MKI	<5 yrs

Computer-aided grading is a rising field and is becoming popular. Many computerized systems are developed for different cancers. Previous work shows grading of neuroblastoma with different methods [7, 8]. In this work support vector machine (SVM) classifier is used for classification which is a popular method in machine learning. It is mainly used to deal with the supervised learning problems. It can be ensured by this method that computational complexity is not increased when non-linear problem is mapped to high dimensional space. It is used to train data with different possible kernel functions which map the training data into kernel space. Further sections include detail about SVM and the attributes used for classification.

Support Vector Machine

Support vector machine is used for classification and regression as represented by Vapnik and his collaborations [9]. It analyses data and recognize patterns. It is defined as systems which uses set of all hypotheses for its processing. It uses high collection of features which are used to characterize our data. This method is learned by an algorithm which helps in training. Theory of machine learning is used for classification of data with the help of this. It maximizes the accuracy of the data. It is a classifier which is derived from learning theory. This learning system thus includes:

- Set of hypotheses
- Collection of many features to classify data
- System is trained by an algorithm

When set of training examples are given, which may belong to different categories, then support vector machine helps in classifying new training examples into different groups. In this examples are represented as points. Set of examples are mapped in this space and then differentiated. New examples are then compared to that space and they are assumed to fall on particular side of the gap they belong to. Support vector machine thus finds a hyperplane. As a result, maximum points of similar type results in same side of the class. Thus hyperplane maximizes the differentiation of two classes. This can be represented below as described in [10].

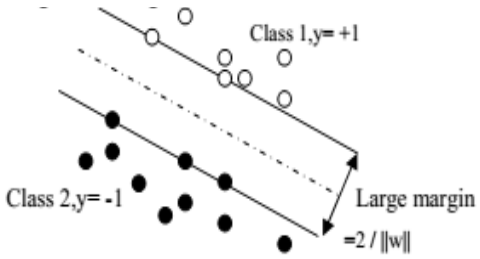


Figure 1: Support vector machine theory

Hence SVM is a useful alternative to neural network and helps in maximizing the area of difference. This helps in reducing of weight vector. Its application includes classification of patient based on the patterns of their brain activity. In the implementation process, we use similar segmentation process with reference [11].

This work is trained by SVM using all available kernel functions. Kernels are used in support vector machine to map the learning data into a high dimensional feature space as described in [12]. Different kernel functions used are linear, quadratic, RBF, polynomial and MLP. Linear kernel is default kernel which is a dot product. It is the simplest kernel function. MLP is a Multilayer Perceptron kernel function. Polynomial kernel functions contains all polynomial terms. It is used when data is normalized.

Polynomial kernel function [13]:

$K(x\,i \cdot x\,j) = [(x\,i \cdot x\,j) + 1]$□

RBF is a Gaussian radial basis function kernel and is used when we don't know much about data.

Radial basis function (RBF) [13]:

$K(x, x\square) = \exp(-\|x - x\square\|^2 / g^2)$

Data Collection

The sample data are collected from National Cancer Institute's (NCI) website [14]. NCI's Office of Cancer Genomics (OCG) provided three programs out of which I have collected dataset from TARGET (Therapeutically Applicable Research to Generate Effective Treatments) data matrix. It is managed by NCI's Office of Cancer Genomics and Cancer Therapy Evaluation Program. Dataset consists of 640 samples consisting of various attributes, out of which required attributes are trained and used for classification. Further section includes the description of the attributes used.

Attributes Description

Dataset consists of various attributes out of which the following attributes are used in this work:

Stage

According to International Neuroblastoma Staging System (INSS), the dataset consists of following stages specified as in [15]:

Stage 1: Cancer is present in one part of the body. It does not spread and can be removed by surgery.

Stage 2a: Tumor is present only in one part that is it is localized but it cannot be removed completely by surgery.

Stage 2b: In this stage also tumor is present in one body part only. It may or may not be removed by surgery. It spreads into nearby lymph nodes.

Stage 3: It is not possible in this stage to remove tumor completely by surgery. Tumor can be large in this stage. Tumor is in one side of the body. It spreads to local areas but not in far apart parts of the body.

Stage 4: It is advanced stage. Tumor in this stage spreads to far apart parts of body like lymph nodes, skin, bone marrow and the other organs.

Stage 4s: It is also called as special neuroblastoma in this stage. It is found in children below one year old. This tumor spreads to liver, bone marrow and skin. All the tumors may be completely removed by surgery in this stage.

MYCN Amplification

MYCN gene status is specified in dataset which is one of factor for classifying Neuroblastoma tumor. It is a protein which controls expressions of many target genes. In dataset this factor is specified by values as: Not Amplified and Amplified [16].

Ploidy
It specifies the number of sets of chromosomes in the nucleus of a cell. In dataset this factor includes the following values: Hyper diploid (DNA Index >1) and Hypo diploid or diploid (DNA Index <=1) [17].

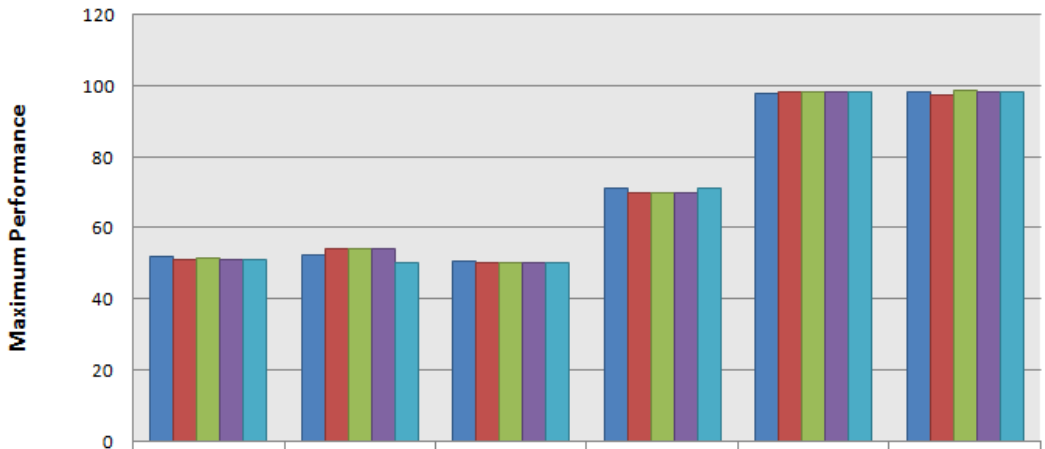
Histology
It has significant impact on prognosis and risk group assignment of neuroblastoma tumor. This dataset describes tumors by INPC as: Favorable and Unfavorable [2]. It is summarized in table 2 and 3.

MKI
MKI is the count of tumor cells present. Mitiosis is the process division of cell into two cells. Karyorrhexis is the division of the nucleus of a dying cell which is damageable in nature [5]. A higher value of MKI shows a higher chance for the tumor to be unfavorable. INPC involves evaluation of tumor by MKI (Mitiosis-karyorrhexis index) of the neuroblastic cells. It has the following values: Low**,** Intermediate, and High.

Diagnostic Category
According to INPC, neuroblastoma is categorized as follows: Neuroblastoma (Schwannianstroma-poor), Ganglioneuroblastoma, intermixed (Schwannianstroma-rich), Ganglioneuroma (Schwannianstroma-dominant), maturing subtype or Ganglioneuroblastoma, well differentiated (Schwannianstroma-rich), Ganglioneuroblastoma, nodular (composite) [4].

Results and Discussions

Result shows that for undifferentiated or poorly differentiated cases, for all features different kernel functions performance is same. MKI and diagnostic category factor shows maximum performance in this case which is 98.12% and 98.44% respectively which implies that these attributes are most responsible for this grade of neuroblastoma. MYCN amplification, stage and ploidy have not much contribution in this case. The approximate occurrence of stage, mycn and ploidy are 51.88%, 54.06% and 50.62% respectively for all kernel functions. Histology attribute has a maximum of 71.25% contribution in this classification using Multilayer Perceptron kernel. It is summarized in figure 2 below.

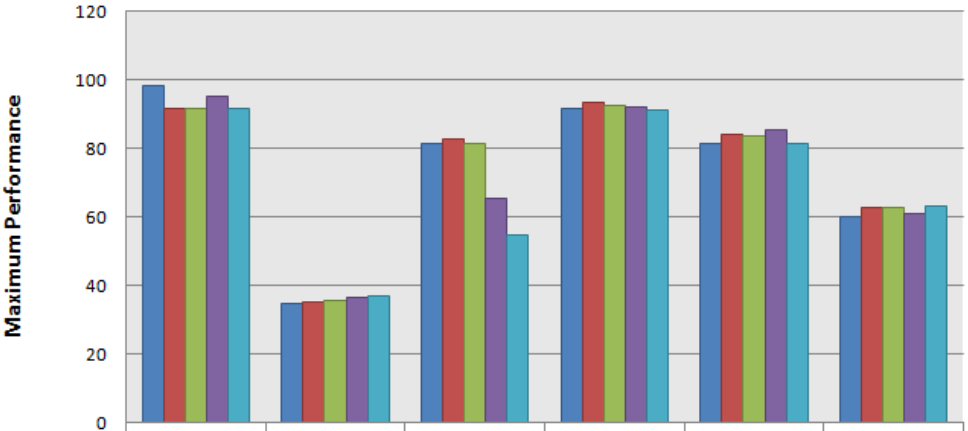


	Stage	MYCN	Ploidy	Hist	MKI	Diag
Linear	51.88	52.5	50.62	70.94	97.81	98.12
Polynomial	51.25	54.06	50.31	70	98.12	97.5
Rbf	51.56	54.06	50	70	98.12	98.44
Quadratic	51.25	54.06	50	69.69	98.12	98.12
Mlp	51.25	50.31	50	71.25	98.12	98.12

Figure 2: Classification rate of undifferentiated or Poorly Differentiating cases

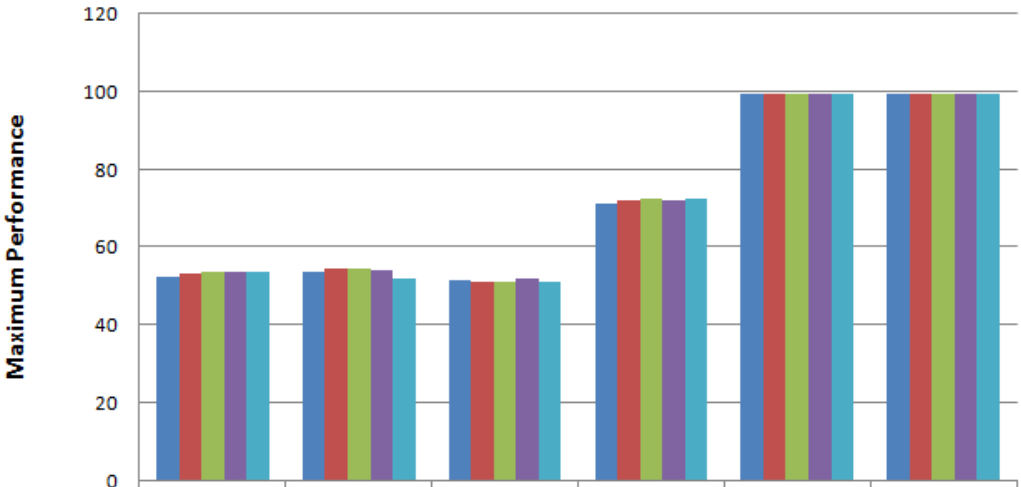
For differentiating cases maximum result is obtained by linear kernel with 98.12% by the stage attribute. It implies that if stage is correctly classified then differentiating cases of neuroblastoma can be classified. Histology attribute is almost equally important for this case with 93.42% efficiency using polynomial kernel. MYCN amplification factor shows least role in this case. MKI shows maximum of 85.27% contribution in differentiated cases. Ploidy attribute has a maximum of 82.76% contribution using polynomial kernel. Observed results are summarized in figure 3.

For classifying unknown cases both MKI and diagnostic category attributes show maximum performance of 99.69%. As in first case, in this case also for each attribute every kernel function performs equally. Stage, MYCN and ploidy factors have less contribution in this case which are 53.92%, 54.55% and 52.04% respectively. Results are shown in figure 4 below.



	Stage	MYCN	Ploidy	Hist	MKI	Diag
Linear	98.12	34.48	81.5	91.54	81.19	60.19
Polynomial	91.54	35.11	82.76	93.42	84.01	62.7
Rbf	91.54	35.42	81.5	92.48	83.7	62.7
Quadratic	94.98	36.36	65.2	92.16	85.27	60.82
Mlp	91.54	36.99	54.55	91.22	81.19	63.32

Figure 3: Classification Rate of Differentiating Cases



	Stage	MYCN	Ploidy	Hist	MKI	Diag
Linear	52.66	53.61	51.72	71.47	99.69	99.37
Polynomial	53.29	54.55	51.41	72.1	99.37	99.37
Rbf	53.61	54.55	51.41	72.41	99.37	99.69
Quadratic	53.92	54.23	52.04	72.1	99.37	99.37
Mlp	53.61	52.04	51.41	72.73	99.37	99.37

Figure 4: Classification Rate of Unknown Cases

Conclusions and Future Work

This work is done for grading of a cancer based on different features. Cancer used in this work is neuroblastoma which is nerve cell cancer. It happens mainly in new born and children less than 3 years of age. Classification is done by Support Vector Machine (SVM) classifier and grading based on International Neuroblastoma Pathology Classification (INPC). The result obtained shows the performance of each attribute taking part in classification. It grades the tumor as poorly differentiated, undifferentiated and differentiating based on the level of mitiosis-karyorrhexis index (MKI), diagnostic category, stage, histology, and ploidy and MYCN status. Each attribute is checked for different kernel functions present in SVM that are linear, quadratic, RBF, polynomial and MLP. From the results it can be concluded that MKI and diagnostic category are important features for undifferentiated or poorly differentiating cases. And stage and histology shows high efficiency while classifying differentiating cases.

In this work each attribute is trained and tested individually by SVM classifier. Future work may include multiple attributes taken together and tested for different grades. Efficiency may be evaluated by taking two or three or all features together. Limitation of this work is that the dataset collected may contain error values.

Acknowledgement

I would like to thank Lovely Professional University for the support on academic studies and providing me such an environment of maintained labs and staff assistance.

I would also like to thank National Cancer Institute for access to dataset on which I performed my training and testing for my work to obtain results and present them.

References

[1] American Cancer Society, http://www.cancer.org (2007)

[2] Shimada, H., et al., 1999, "The International Neuroblastoma Pathology Classification", Cancer, Vol. 86, No. 2, pp. 364-372.

[3] Brodeur, G.M., et al., 1984, "Amplification of N-myc in untreated human neuroblastomas correlates with advanced disease stage", Science, Vol. 224, No. 4653, pp. 1121-1124.

[4] Shimada, H., et al., 1999, "Terminology and morphologic criteria of neuroblastic tumors", Cancer, Vol. 86, No. 2, pp. 349-363.

[5] Sertel, O., Catalyurek, U.V., Shimanda, H., and Gurcan, M.N., 2009, "Computer-aided Prognosis of Neuroblastoma: Detection of mitiosis and karyorrhexis cells in digitized histological images", 31st Annual International Conference of the IEEE EMBS, pp. 1433-1436.

[6] Shimada, H., et al., 2001, "International Neuroblastoma Pathology Classification for Prognostic Evaluation of Patients with Peripheral Neuroblastic Tumors", Cancer, Vol. 92, No. 9, pp. 2451-2461.

[7] Gurcan, M. N., Pan, T., Shimanda, H., and Saltz, J., 2006, "Image Analysis for neuroblastoma classification: Segmentation of cell nuclei", IEEE Int. Conf. of Eng. In Medicine and Biology, New York City.

[8] Saxena, P., and Singh, S.K., 2012, "Noble approach for texture classification of H&E stained histopathological image by Gaussian wavelet", IEEE 12th International Conference on Intelligent Systems Design and Application (ISDA), pp. 375 – 379.

[9] Vapnik, V., 1998, "The nature of statistical learning theory" New York: Springer Verlag.

[10] Wang, X., Wang, S., Zhu, Y., and Meng, X., 2012, "Image segmentation based on Support Vector Machine" IEEE 2nd International Conference on Computer Science and Network Technology (ICCSNT), pp. 202 – 206.

[11] Mangasarian, O. L., and David R., 2000, "Active set support vector machine classification" Neural Information Processing System 2000 (NIPS).

[12] Smits, G.F., and Jordaan, E.M., 2002, "Improved SVM regression using mixtures of kernels", International Joint Conference on Neural Networks, pp. 2785-2790.

[13] Men, H., et al., 2009, "Study on Classification Method Based on Support Vector Machine", First International Workshop on Education Technology and Computer Science, Vol. 2, pp. 369-373.

[14] National Cancer Institute, http://www.ocg.cancer.gov

[15] Monclair, T., et al., 2009, "The International Neuroblastoma Risk Group (INRG) Staging System: An INRG Task Force Report", Journal of clinical oncology, Vol. 27, No. 2, pp. 298-303.

[16] Spitz, R., et al., 2006, "MYCN-status in Neuroblastoma: characteristics of tumors showing amplification, gain, and non-amplification", European Journal of Cancer, Vol. 40, Issue 18, pp. 2753-2759.

[17] Eckschlager, T., et al., 1999, "DNA ploidy in neuroblastoma", Neoplasma, pp. 23-26.