

Classification of Follicular Lymphoma Grades Using Support Vector Machine

KirandeepKaur¹, Sanjay Kumar Singh¹, PrateekAgrawal¹, Anjali Goyal²
*¹Department of Computer Science Engineering, Lovely Professional University,
Phagwara, Punjab, India
Email: {kirankambo18; sanjayksingh.012; prateek061186}@gmail.com*
*²Dept. of Computer Applications, GNIMT, Ludhiana
Email: anjali.garg73@gmail.com*

Abstract

In this paper, an approach for grading of follicular lymphoma has been introduced which classifies the grade according to the attributes taken as input, which are stage, stage group, tumor size and performance status. The grading tells how fast the lymphoma is spreading in the patient's body which is important for doctors in treatment of this disease. This proposed system performs grading by using support vector machine for classification with different kernels which are polynomial, multi-layer perceptron (mlp), linear, quadratic and radial basis function (rbf), which classifies the data into respective grades of follicular lymphoma, which are Grade 1, Grade 2, Grade 3A and Grade 3B.

Index Terms- Follicular Lymphoma, stage, stage group, tumor size, performance status.

Introduction

Follicular lymphoma (FL) is the cancer of lymphatic system which is the most widespread category of Non-Hodgkin lymphoma. It is B-cell type of lymphoma in which B-lymphocytes develop in circular pattern and these lymphoma cells group together in clusters in a lymph node or other tissues to form a tumor. These enlarged irregular B-cells are called centroblasts. This lymphoma is diagnosed by examining the shapes of centroblasts under microscope.

It is basically slow-growing type of lymphoma, takes several years to develop in the body. It typically occur in adults with age 60 years and older and is not commonly found in young people. Sometimes follicular lymphoma transforms into a fast-growing (aggressive) diffuse large B-cell lymphoma (DLBCL). This lymphoma is slightly more commonly found in women than men and found very less in Asians and

Blacks as compared to people of other ethnicities. This disease is not inherited one. Children of the patient with this disease do not have any threat of developing it. Some of the symptoms are swelling of lymph nodes mainly in neck, groin areas and armpits, high fever, high sweating at night, sudden weight loss, itching, fatigue, loss of appetite, difficulty in breathing etc. So it is very essential to diagnose the disease at an earlier stage to save the person from loss of life.

To examine the grade of follicular lymphoma it is essential to count the number of centroblasts / high power microscopic field (HPF). As there can be misinterpretations in counting centroblasts manually by the doctors and also it consumes a lot of time to manually interpret slides from microscope, we are proposing a system which will use support vector machine to classify the data and thus will perform grading. According to World Health Organisation (WHO) the grades have been classified into 3 broad categories: Grade I has 0 to 5 centroblasts / HPF, Grade II has 6 to 15 centroblasts / HPF and Grade III has more than 15 centroblasts / HPF.[1] The first 2 categories are low grade and the third category is high grade cancer. Further, Grade III is sub-categorised as Grade 3A which is high grade lymphoma , Grade 3B which is called Diffuse Large B-cell Lymphoma (DLBL).It is very important for doctors to know the grade in order to treat the patient. Depending on the grade particular type of treatment is given to the patient so as to save his/her life. Thus it is very important to correctly classify the lymphoma grade.

Table 1: WHO Criteria Of Follicular Lymphoma Grades[2]

Grade	Centroblasts
Grade 1	0 to 5 centroblasts / HPF
Grade 2	6 to 15 centroblasts / HPF
Grade 3	More than 15 centroblasts / HPF
Grade 3A	Centrocytes still present
Grade 3B	The follicles consist almost entirely of ccentroblasts

Where hpf is high-power field which is the area to be observed specifying the magnification of lens used.

Table 1.2: Architecture Patterns in Follicular Lymphoma[3]

Follicular Lymphoma Pattern	Architecture
Follicular	more than 75% follicular
Follicular and diffuse	25 to 75% follicular
Focally follicular	less than 25% follicular
Diffuse	0% follicular

Data Collection

The dataset has been taken online from Office of Cancer Genomics, National Cancer Institute under access of CGCI (Cancer Genome Characterization Initiative) Data Matrix Clinical File from its website.[11] This dataset has multiple attributes related to follicular lymphoma having a large set of values. Out of those values and attributes, we have taken 4 attributes and 110 instances, that is, clinical report of 110 patients.

Input Data Attributes

The input data that has been taken has 4 attributes which are used to perform grading of follicular lymphoma. These 4 attributes are : stage, stage group, tumor size and performance status. These attributes are useful for the diagnosis of follicular lymphoma Grade. Following is the description of input data attributes:

A. Stage

Stage tells that to where and which body parts the lymphoma has spread. It specifies that which organs have been affected. There are 4 stages:

- Stage 1: Only one lymph node region (neck, armpit, groin or chest) or only one lymph structure (lymph nodes, spleen, and thymus gland) is affected.
- Stage 2: lymphoma has spread to two or more regions or structures of the lymph nodes on one side of diaphragm.
- Stage 3: lymphoma has spread in structures or regions of lymph nodes on both of the diaphragm sides.
- Stage 4: lymphoma has spread to many organs or parts of the body or tissues like bone marrow

While assigning a stage, one letter is also included, A or B, which denotes some symptoms fever, loss in weight or sweating at night etc. "A" tells that patient has such symptoms; "B" tells that patient does not have such symptoms. Sometimes E is also included which tells that lymphoma has spread to organs outside the lymph nodes.[6]

B. Stage Group

Stages have been categorized under two groups:

- Advanced: This group tells that patient has lymphoma on both sides of diaphragm and size of tumor is more than 10 cm. The patient has B symptoms (night sweats, etc.).
- Limited: This group tells that the patient has lymphoma on only one side of diaphragm and tumor size is small. The patient has no B symptoms.[7]

C. Tumor Size

Size of tumor is important parameter to classify the grades. This parameter tells how much mass of tumor has developed in the affected area. This size may vary depending upon the grade and stage of lymphoma. The tumor size here is taken in millimeters (mm).

D. Performance Status

Performance status is useful in telling the physical condition of the patient, which can be used as a parameter to analyze the grade. WHO performance status is:

- 0: The person is completely energetic and can do all kind of work without restriction.
- 1: The person is limited to physically tiring activities but ambulatory and they can perform light activities in work.
- 2: Ambulatory and can take care of his/herself but cannot perform any activity at work.
- 3: The person is restrained in chair or bed more than 50% of his/her working hours.
- 4: Totally immobilized. Cannot take care of him / herself. Totally restrained to chair or bed.[8]

Support Vector Machine

Classification is done for classifying the objects/features in the image. SVM will classify the data and thus grading will be performed based on this classification. It is based on supervised learning in which some training examples are presented, each having its own category, SVM allocate new input examples which we have not presented to machine before, into their respective categories. It builds a hyper plane such that the separation margin between the two kind of examples is maximized and separates the positive examples from negative examples. SVM is easy to use, performance is good, and it can solve large number of problems with same algorithm with less tuning.

The margin of separation is formed by optimal separation in hyperplane between the set of classes of problem. The importance of optimal solution results to lower probability of misclassification when separating training examples accordingly.[4]

Here, the classification has been done using five different kernels namely, linear, quadratic, radial basis function (rbf), polynomial and multi-layer perceptron (mlp). The results have been discussed that with which attributes and kernel we obtained maximum performance. The basic idea of support vector machine has been shown in the figure 1.

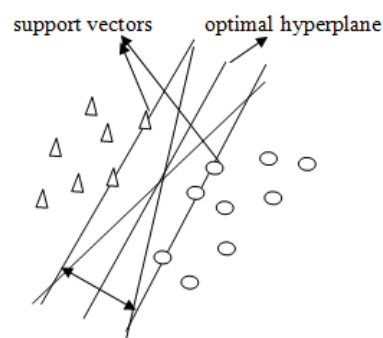


Figure 1: Idea of optimal hyperplane for linearly separable patterns

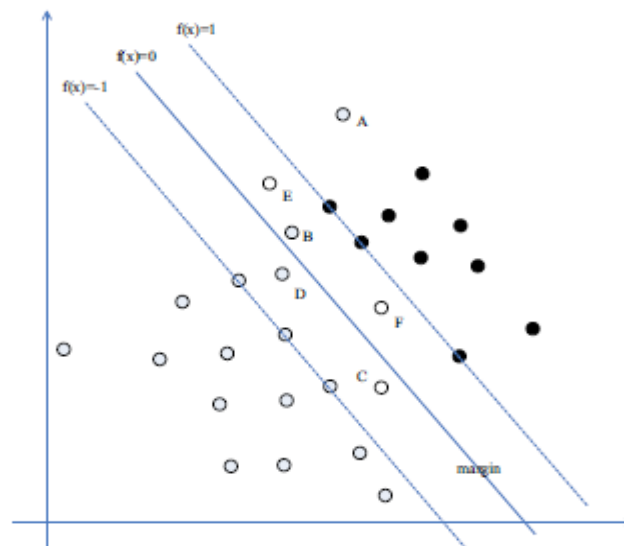


Figure 2: SVM for non-linear separable input patterns. A,B, E and F are misclassified samples[5]

Using different kernel function $K(x_i, x_j)$ can construct different types of learning machines with non-linear decision-making side on the input space, and lead to different support vector algorithms.[9]

Polynomial function is a d-order classifier.[9]

Polynomial kernel function:

$$K(x \square \cdot x_j) = [(x \square \cdot x_j) + 1] \square [10]$$

Radial basis function:

$$K(x, x \square) = \exp(-\|x - x \square\|^2 / g^2) [10]$$

Grading

The final output is the grading of follicular lymphoma which tells the grade of follicular lymphoma that the patient is having. Grade specifies at what speed the lymphoma is spreading in the body. The output is in the form of Grade 1, Grade 2, Grade 3A and Grade 3B.

Results and Discussion

The result for every Grade has been shown in the form of bar graphs for every attribute and kernel function. The results have been taken using MATLAB. It is clearly shown that Stage attribute has been the best which gave the maximum performance for every Grade under polynomial kernel on average.

Grade 1 has been best classified with polynomial kernel with 81.82% of maximum performance and stage attribute has given the best classification as shown in figure 2.

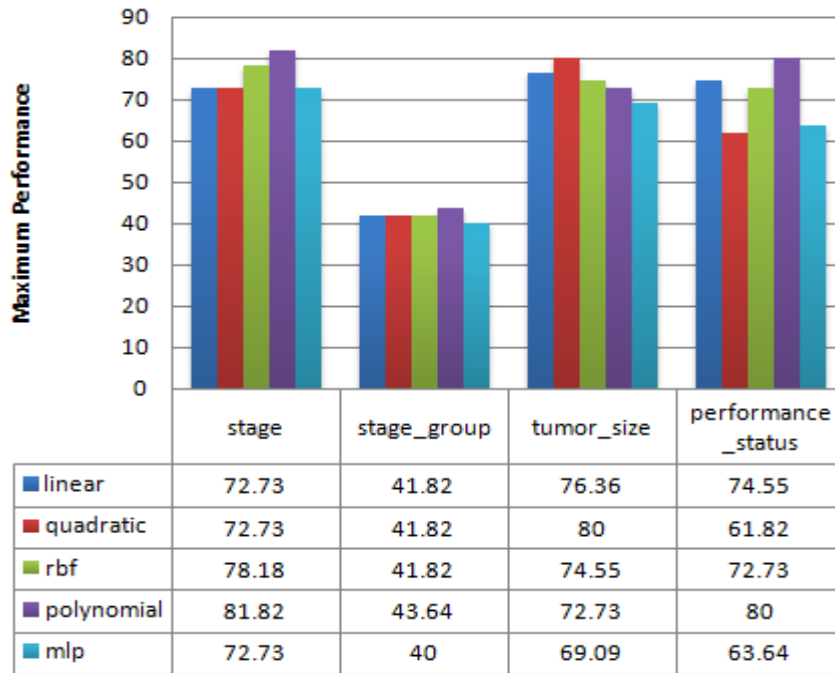


Figure 2: Classification rate of Grade 1

Grade 2 has been best classified with rbf kernel with 87.27% of maximum performance and stage attribute has given the best classification as shown in figure 3.

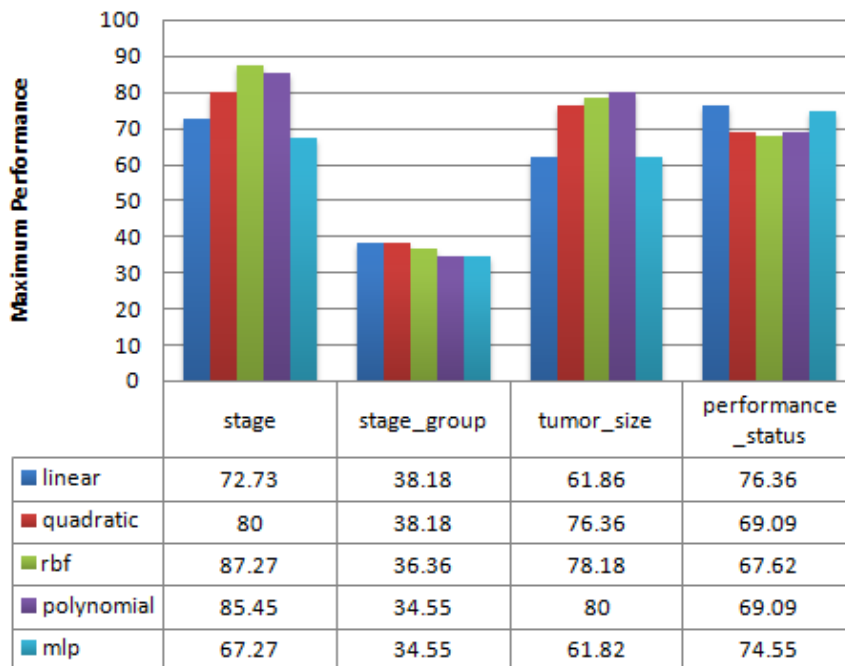


Figure 3: Classification rate of Grade 2

Grade 3A has been best classified with quadratic and polynomial kernel with 96.30% of maximum performance, tumor size and stage attribute respectively have given the best classification as shown in figure 4.

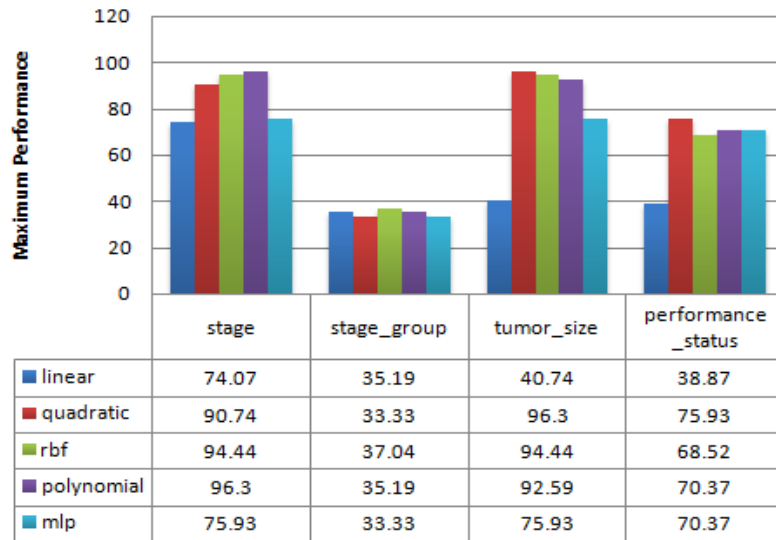


Figure 4: Classification rate of Grade 3A

Grade 3B has been best classified with polynomial kernel with 81.48% of maximum performance and stage attribute has given the best classification as shown in figure 5.

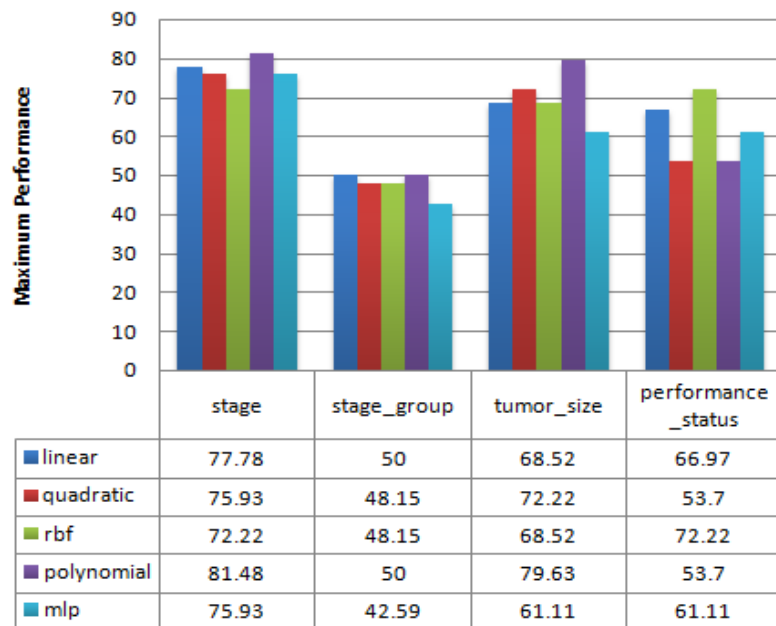


Figure 5: Classification rate of Grade 3B

Acknowledgement

I would like to present my deepest gratitude to the members of Office of Cancer Genomics, National Cancer Institute from where the data has been taken.

I would also like to thank Lovely Professional University, which has provided us with labs for practical implementations and helped me with my research work.

Conclusion and Future Work

This proposed system has been designed to perform grading of follicular lymphoma in lesser time and to give accurate results. It can be used to diagnose the follicular lymphoma that how fast the lymphoma is spreading in the body. This system has classified the data into grades by using 4 different attributes and 5 kernels of support vector machine and result has been analyzed for each attribute corresponding to each kernel.

We concluded that stage attribute has given best classification with 81.82% of maximum performance with polynomial kernel for Grade 1, 87.27% with rbf kernel for Grade 2, 96.30% with polynomial kernel for Grade 3A and 81.48% with polynomial kernel for Grade 3B. Also we concluded that with polynomial kernel function, best classification results have been obtained.

Since this system has classified the data into Grades by using single attribute at a time. The Future Work of this system will be classification of Grades using combination of multiple attributes. The limitation of this system is that the data that we have taken online might have some mistaken values.

References

- [1] Sertel, O., Kong J., Catalyurek, U. V., Lozanski, G., Saltz, J., H., and Gurcan M., N., 2008, "*Histopathological Image Analysis Using Model-Based Intermediate Representations and Color Texture: Follicular Lymphoma Grading*", Journal of Signal Processing Systems, vol. 55, Issue 1-3, pp. 169-183.
- [2] Saxena, P., Singh, S., K., and Agrawal, P., 2013, "*Texture Classification of Biased Cytoplasmic Tissue Sample from Histopathological Imagery by Gabor Application*", Journal of Network and Innovative Computing ISSN 2160-2174 Volume 1, pp. 248-259.
- [3] McNamara, C., Davies, J., Dyer, M., Hoskin, P., Lyttelton, M., Marcus, R., Montoto, S., Ramsay, A., and Ardesna, K., 2011 "*Guidelines on the investigation and management of follicular lymphoma*", 2011 Blackwell Publishing Ltd, British Journal of Haematology, 156, 446–467.
- [4] Nooritawati, M., T, and ShahRizam M., S., B., 2012, "*A Comparative Study in Kernel-Based Support Vector Machine of Oil Palm Leaves Nutrient Disease*", International Symposium on Robotics and Intelligent Sensors, Procedia Engineering ,41,1353 – 1359.

- [5] Fa Zhu, Ning Ye, Sheng Xu and XiaojunGu, 2011 “*Support Vectors Classification and Incremental Learning*”, Information Technology and Artificial Intelligence Conference, vol. 1, pp. 206-210.
- [6] <http://www.uptodate.com/contents/follicular-lymphoma-in-adults-beyond-the-basics>
- [7] <http://www.cancerresearchuk.org/about-cancer/type/non-hodgkins-lymphoma/treatment/staging-non-hodgkins-lymphoma>
- [8] WSW Shalaby - Manual of Gynecologic Oncology, 2011
- [9] Hong, S., Yan, K., Hong, Z., 2010 “*Chest DR Image Classification based on SupportVector Machine*”, Second International Workshop on Education Technology and Computer Science, vol. 1, pp. 170-173.
- [10] 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.
- [11] National Cancer Institute , <https://ocg.cancer.gov/programs/cgci/projects>.

20450

KirandeepKaur