

Detecting Virulent Cells of *Cryptococcus Neoformans* Yeast: Clustering Experiments

Jinshuo Liu^{1,2}, Peter van der Putten², Ferry Hagen³, Xinmeng¹ Chen, Teun Boekhout³,
Fons J. Verbeek²

jsureliu@gmail.com, putten@liacs.nl, xinmeng@whu.edu.cn
{hagen,boekhout}@cbs.knaw.nl, fverbeek@liacs.nl

1: Computer School, Wuhan University, Wuhan, P.R.China

2: Leiden Institute of Advanced Computer Science, Leiden University, Leiden, The Netherlands

3: CBS Fungal Diversity Institute, Netherlands Academy of Science and Arts, Utrecht, The Netherlands

Abstract — The yeast *Cryptococcus neoformans* can cause dangerous infections such as meningitis. The presence of a thick capsule is shown to be correlated with virulence of a yeast cell. This paper reports on our approach towards developing a classifier for detecting virulent cells in images. We present our methods for creating samples, collecting images, preprocessing the images, identifying cells and creating features for each cell. Unsupervised clustering experiments have provided preliminary evidence that our methods results in features that can successfully be used to group and distinguish virulent from normal cells. In our future work we plan to use the same methods and feature set to build supervised classification models.

I. INTRODUCTION

The basidiomycetous pathogenic yeast *Cryptococcus neoformans* can cause meningitis, meningoencephalitis, and pulmonary and skin infections. Infections occur mainly in immunocompromized patients, i.e. HIV-infected patients, transplantation patients and leukemia patients [3]. One of the most significant virulence factors of the fungus is the presence of an extra cellular polysaccharide capsule [2,6,8,11,12].

Complementation of a capsule-deficient mutant clearly showed the relation between the presence of a capsule and virulence [4]. As the thickness of this capsule can vary between strains, specific genetic constructs related to capsule biosynthesis, and between different environmental conditions [1,3,6,8,15,18] there is a need for easy-to-perform, automated image analysis tools to detect pathogen yeast by analyzing the capsule thickness of cryptococcal cells.

In this paper we present initial results towards building such a classifier. We present our methods for creating samples, collecting images and extracting features. To explore the validity of our approach we performed unsupervised clustering on the generated data set. Using supervised algorithms to build a classifier will be the next step but is out of scope for this paper.

The data set contains two classes, namely those with a thick(er) capsule (i.e grown without NaCl, see below) and those with a thin(ner) capsule (i.e. grown with the addition of 1M NaCl, see below), thus simulating the different capsule morphologies that may occur in different organs, growth conditions etc.

The remaining part of this paper is organized as follows: In section 2 we present our methods for image collection, preprocessing and feature extraction. In section 3 the results of the clustering experiments will be given. We conclude the paper in section 4.

II. IMAGE COLLECTION, PREPROCESSING AND FEATURE EXTRACTION

Samples and images were provided by the CBS Fungal Diversity Institute (Utrecht, the Netherlands). We identify individual cells in images and extract a set of features specifically tuned towards the classification problem.

A. Image Collection

As most analyses of cryptococcal capsules are being made using negatively stained cells with India ink, we used this staining method for our experiments. The capsulate strains CBS 950 (*C. neoformans*) and CBS 919 (*C. gattii*) were grown for 48 h at 25°C on YPGA (1 % yeast extract-0.5 % peptone-4 % glucose) medium without and with the addition of 1M NaCl, the latter with the purpose to suppress capsule development [6]. Capsular size was investigated by suspending cells in India ink, and photographs were taken with a Zeiss Axioskop microscope using phase-contrast optics using a 100 x Plan-neofluar 100x/1.30 objective. Images were taken using a ADIMEC MX12P digital camera and stored as tiff files.

B. Image preprocessing phase

In some of the samples yeast cell images are difficult to interpret and therefore some preprocessing of the images was required to improve the quality of the

images make the feature extraction phase more reliable. Preprocessing is always necessary whenever the data to be mined is noisy, inconsistent or incomplete. Preprocessing significantly improves the effectiveness of the data mining techniques.

All objects, i.e. yeast cells, on the image border were removed from the analysis. The application of a Kuwahara filter with round mask [7,19] was sufficient for enhancement and noise suppression. Moreover, the objects in the image that were too small were removed because we believe that those yeast cells are not mature enough to analyze their biology.

Our method of image segmentation [15] is to identify the yeast cell as a separate object. For each yeast cell the capsule (outer bound) of the cell is detected and the cell is divided into inner part and the outer part. Additionally, our image segmentation method considers clumps of cells and divides these into separate objects using a watershed segmentation approach. The results are evaluated in a heuristic so that all cells are mature cells; budding cells are discarded [2]. The segmentation of the capsule was accomplished with region growing method starting from four corners of the bounding box of the cell.

The resulting collection of images consists of 28 thick capsule cells and 44 thin capsule cells.

C. Feature Extraction

After preprocessing and segmentation features relating to the classification are extracted from the objects in the images. Features we selected to use are moments and the moments invariants [9,10,13] that are derived from the moments. Some other features we selected can be interpreted directly by a fungal biologist (have meaning to how they would observe the yeast cell), i.e. ratio of radius, ratio of area, ratio of intensity of capsule and inner part. In addition, moments can provide unique shape characteristics of an object [9].

The features can be characterized in two groups: i.e.: Interdependent attributes consisting of visible features (area, diameter and intensity for the binary and gray value images.) and statistical features [10]. Derived attributes: The attributes are obtained from computation with results from the interdependent attributes.

The moment of order $(p+q)$ in a gray-level image are described by:

$$m_{pq} = \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} x^p y^q f(x, y) dx dy \quad (1)$$

From the moments the centralized moments and the normalized moments are derived. The normalized moments are the basis for the 7 moment invariants

[9].

Rather than using the moments of the image as a whole we have applied these to the objects (yeast cell) that were extracted from the images in the segmentation process [16].

The statistical features: major variance, skewness, kurtosis, gyration are derived from first order, second order and third order moments. We use standard measure for variance, skewness and kurtosis along an axis [16]. For completeness, skewness is defined as:

$$S_k = \frac{1}{N} \left(\frac{(x-\bar{x})^3}{\sigma} \right) \quad (2)$$

and kurtosis as:

$$Kurt = \frac{1}{N} \left(\frac{(x-\bar{x})^4}{\sigma} \right) - 3 \quad (3)$$

All these extracted features are computed for each of objects in which each object is also used as a binary mask for the original image to obtain density features. So both binary images and gray value images are used in the feature extraction.

III. MINING & INTERPRETATION

In preparation for building a supervised classifier and to validate our preprocessing approach we have carried out explorative unsupervised clustering experiments. We have used a variety of clustering techniques, including simple K-Means, Expectation Maximization, COBWEB and FarthestFirst clustering, using the WEKA data mining tool [17].

The implementation of COBWEB did not allow us to control the number of cluster directly and FarthestFirst resulted in unbalanced clusters in terms of the number of instances, so EM and K-Means gave the best results for this problem.

The clustering was done on a subset of features that were found to be predictive on a related set of images (same collection but nigrosine stained). Examples of features selected were 'area': ratio of area of capsule and inner part; 'alfa': ratio of semiminor of capsule and inner part; 'beta': ratio of semimajor of capsule and inner part; 'diameter difference': ratio of radius of capsule and inner part; the gray level and binary image $Mi(0)$ and $Mi(1)$ and the mean intensity ratio.

In Figure 1 we give an example of ten cells, four of which fall into cluster 0 and have thick capsules, six fall into cluster 1 and have thin capsules. This clustering was generated using k-Means clustering (random seed 10). In Table 1 we provide an overview of the main features values for these cells and averages and standard deviations of the cluster centers. It can be clearly be seen that in this example the two different types of cells are distinguished well.

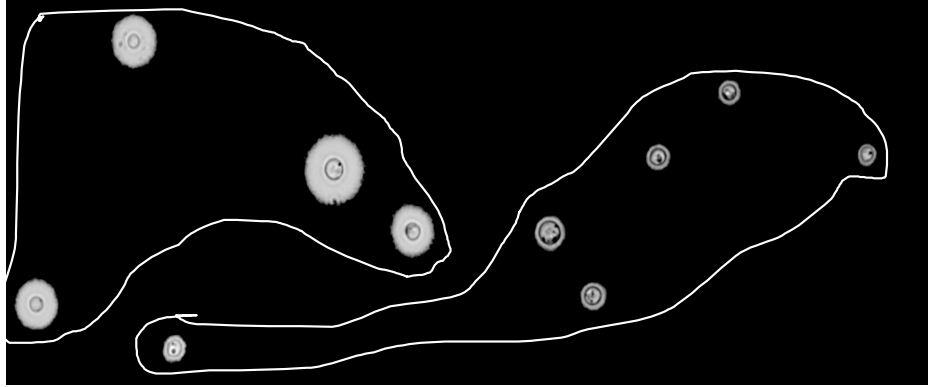


Figure 1. Four cells on the left fall into cluster one (thick capsules corresponding to virulent cells), the six on the right in cluster two (thin capsules corresponding to normal cells)

	cluster		Area ratio	Alfa ratio	Beta Ratio	Diadiff Ratio	Integreted Intensity Ratio	Gray C/I Mi[0]	Gray C/I Mi[1]	Binary C/I Mi[0]	Binary C/I Mi[1]	Mean Intensity Ratio
		μ	0.8779	0.5433	0.5432	0.8885	0.7967	429	30.5909	2.7535	12.7	1.1122
		σ	0.1547	0.0191	0.0276	0.0891	0.1631	67.388	56.5388	56.5388	/	0.1087
1	Cluster0		0.9481	0.5425	0.5512	0.9615	0.9246	4.88e+02	2.42e+02	2.93e+00	1.27e+01	1.03e+00
2			0.7346	0.5424	0.5194	0.8053	0.6485	3.39e+02	9.03e-01	2.46e+00	1.41e-01	1.13e+00
3			0.9315	0.5312	0.5142	0.9007	0.8113	3.389e+02	1.62e+00	2.85e+00	5.35e+00	1.15e+00
4			0.8287	0.5562	0.5512	0.8628	0.7443	3.82e+02	3.71e+01	2.65e+00	4.33e+01	1.11e+00
		μ	0.3238	0.4669	0.4511	0.4655	0.2383	104.3059	4.2602	4.2602	6.58	1.3905
		σ	0.1622	0.1117	0.126	0.149	0.126	67.5845	16.1438	18.1438	/	0.2736
5	Cluster1		0.3152	0.4561	0.4491	0.4511	0.2034	1.92e+02	2.82e-01	1.64e+00	6.58e+00	1.55e+00
6			0.3555	0.4907	0.4467	0.4915	0.2416	2.13e+02	2.12e-02	1.70e+00	4.12e-02	1.47e+00
7			0.2367	0.4218	0.4149	0.4030	0.1624	1.84e+02	2.39e-01	1.48e+00	1.79e+00	1.46e+00
8			0.4751	0.5117	0.4727	0.5624	0.3163	2.49e+02	1.17e-01	1.97e+00	8.56e-01	1.50e+00
9			0.4734	0.5279	0.4555	0.5441	0.2960	2.33e+02	1.06e-01	1.93e+00	6.43e-01	1.60e+00
10			0.2683	0.4246	0.4107	0.5081	0.2581	2.96e+02	4.18e-01	1.54e+00	6.94e-01	1.04e+00

Table 1. Cluster means, standard deviations and attribute values of cells from figure 1.

IV. CONCLUSIONS

The clustering experiments have provided preliminary evidence that our methods for image collection, preprocessing, identification of individual cells and feature extraction result in features that can successfully be used to group and distinguish virulent from normal cells. In our future work we plan to use the same methods and feature set to build supervised classification models.

REFERENCES

- [1] I. Bose, A. J. Reese, J. J. Ory, G. Janbon, and T. L. Doering, "A yeast under cover: the capsule of *Cryptococcus neoformans*," *Eukaryot Cell*, vol. 2, pp. 655-63, 2003.
- [2] G. S. Bullmer, M. D. Sans, & C. M. Gum, "*Cryptococcus neoformans*," *Journal of Bacteriology*, No. 1967, p.1475-1479.
- [3] Casadevall A, Perfect JR 1998. *Cryptococcus neoformans*. ASM Press, Washington, D.C., USA, pp. 1-541.
- [4] Chang YC, Kwon-Chung, KJ. 1994. Complementation of a capsule-deficiency mutation of *Cryptococcus neoformans* restores its virulence. *Mol. Cell. Biol.* 14: 4912-4919.
- [5] R. O. Duda, P. E. Hart, D. G. Stork (2001) *Pattern Classification*, 2nd Ed. Wiley & Sons, Inc.
- [6] Dykstra MA, Friedman, L, Murphy JW, 1977. Capsule size of *Cryptococcus neoformans*: control and relationship to virulence. *Infect. Immun.* 16:129-135.
- [7] R. C. Gonzales and R. E. Woods, *Digital Image Processing*, 1993.
- [8] G. Janbon, "Cryptococcus neoformans capsule biosynthesis and regulation," *FEMS Yeast Res* 4, 765-771 (2004).
- [9] M. Hu, "Visual pattern recognition by moment invariants," *IRE Trans. Inf. Theory* IT-8, 179-187 (1962).
- [10] L. Keyes, AC. Winstanley, "Using moment invariants for classifying shapes on large scale maps," *GIS Research UK*, York, April 2000.

- [11] M. L. Littman, "Capsule synthesis by *Cryptococcus neoformans*," *Trans N Y Acad Sci* 20, 623-48 (1958).
- [12] M. L. Littman, & E. Tsubura, "Effect of degree of encapsulation upon virulence of *Cryptococcus neoformans*," *Proc Soc Exp Biol Med* 101, 773-7 (1959).
- [13] F. Mindru, T. Tuytelaars, L. V. Gool, T. Moons, "Moment invariants for recognition under changing viewpoint and illumination," *Computer Vision and Image Understanding* 94, 3-27 (2004).
- [14] J. Rivera, M. Feldmesser, M. Cammer, and A. Casadevall, "Organ-dependent variation of capsule thickness in *Cryptococcus neoformans* during experimental murine infection," *Infect Immun*, vol. 66, pp. 5027-30, 1998.
- [15] M. Sonka, V. Hlavac, and R. Boyle, "Image processing, analysis and machine vision," United states of America: Brooks/ Cole, a division of Thomson Asia Pte Ltd.
- [16] F. J. Verbeek, "Three Dimensional reconstruction of biological objects from serial sections including deformation correction," (1995, Delft) pp.87-93.
- [17] I. H. Witten and E. Frank, "Data mining: practical machine learning tools and techniques with Java implementations," Morgan Kaufmann Publishers, 1999.
- [18] O. Zaragoza, A. Casadevall, "Experimental modulation of capsule size in *Cryptococcus neoformans*," *Biological Procedures Online* • Vol. 6 No. 1 March 3, 2004, www.biologicalprocedures.com
- [19] http://www.fch.vutbr.cz/lectures/imagesci/includes/harfa_screenshots_smoothkuwahara.inc.php

