Challenges of Trustable AI and Added-Value on Health
B. Séroussi et al. (Eds.)

© 2022 European Federation for Medical Informatics (EFMI) and IOS Press.
This article is published online with Open Access by IOS Press and distributed under the terms of the Creative Commons Attribution Non-Commercial License 4.0 (CC BY-NC 4.0).
doi:10.3233/SHTI220626

Differentiation of Cell Painted Organelles Using Non Local Texture Descriptor and Random Forest Approach

Sreelekshmi PALLIYIL SREEKUMAR [a,1], Rohini PALANISAMY[b] and Ramakrishnan SWAMINATHAN[a]
[a] *Biomedical Engineering Group, Centre of Excellence in Medical Device Regulations and Standards, Department of Applied Mechanics, Indian Institute of Technology Madras, Chennai, Tamil Nadu, India*
[b] *Indian Institute of Information Technology, Design and Manufacturing, Kancheepuram, Chennai, Tamil Nadu, India*

Abstract. Discriminating the cell organelles from microscopic images is a challenging task due to their high similarity in image appearance. In this work, an attempt has been made to differentiate nuclei, Endoplasmic Reticulum (ER) and cytoplasm using a texture pattern descriptor and Random Forest classifier. For this, Cell Painted public dataset from Broad Bioimage Benchmark collection are considered. Texture features are extracted from each image using Non Local Binary Pattern (NLBP) that captures the relationship between global pixels and sampling instances in a local neighborhood. Non local central pixels called anchors are derived from central pixels of image patches and compared with sampling instances. Binary string generated from this is encoded into 29 patterns. Statistical one-way analysis of variance (ANOVA) is performed to select significant features and are validated using Random Forest classifier. The dependency of classifier performance on the local patch radius (R) and the number of anchors (K) are also evaluated. The results indicate that 8 patterns out of 29 are showing strong inter class variability with high F value. Classification accuracy of 84% is achieved with R=3 and K=5. Experimental results demonstrate that the proposed work captures complex patterns in cell structure useful for differentiating cell components which can be employed for evaluating the cytotoxic effects in cell lines.

Keywords. Cell Painting, Non Local Binary Pattern, Texture Feature, Random Forest

1. Introduction

Cell Painting is an advanced imaging technique that uses different fluorescent probes to target specific cell organelles such as nuclei, Endoplasmic Reticulum (ER), cytoplasm, mitochondria and Golgi apparatus for profiling subtle patterns in cell structure [1]. From the previous studies it can be seen that textures of different cell components are different

[1] Corresponding Author, Sreelekshmi Palliyil Sreekumar, Biomedical Engineering Group, Centre of Excellence in Medical Device Regulations and Standards, Department of Applied Mechanics, Indian Institute of Technology Madras, Chennai, Tamil Nadu, India; E-mail: lekshmishyam2009@gmail.com.

[2]. Hence the spatial relationship in the Cell Painted microscopic images can be effectively measured by texture patterns.

Local binary descriptors have received wide acceptance in characterizing the local features in an image [3]. Local binary pattern (LBP) and its derivatives focus on the spatial relation between central pixel and sampling instances in the local neighborhood [4]. These methods are inadequate in describing the long range pixel interaction that occurs outside a compact region which can also be considered as important for feature representation. In this work, wide range pixel relationship is captured by Non Local Binary Pattern (NLBP) based on global image statistics rather than local connected region in the image [5].

2. Methods

The proposed methodology for categorizing the cell organelles is described in Figure1.

2.1 Image Dataset

For this study, Cell Painted images of Human U2OS cells are obtained from Broad Bioimage Benchmark Collection [6]. 3456 images of nuclei,3456 images of ER and 3456 images of cytoplasm in 16-bit grayscale of size 1080x1080 pixels are considered.

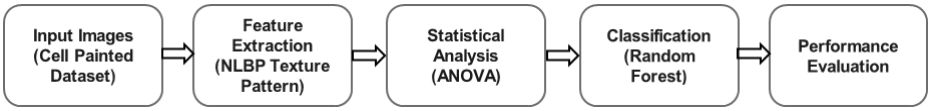


Figure 1. Block diagram of proposed Methodology

2.2 Non Local Binary Pattern

To extract texture features, a local image patch of size ω x ω with central pixel x is considered and their gray values are sorted in ascending order as given in the Eq. (1)

$$\bar{g}'_{c1}, \dots\dots, \bar{g}'_{cN} := \text{sort}(\bar{g}'_{c1}, \dots\dots, \bar{g}'_{cN}) \quad (1)$$

where $\bar{g}'_{cN}$ represents the gray value of the N^{th} sorted central pixel and N is the total number of central pixels [5]. The sorted pixels are then divided into K equal intervals and the anchors are calculated for each interval using Eq. (2)

$$g_{Ak} = \frac{1}{\lfloor N/K \rfloor} \sum_{n=(k-1)\lfloor N/K \rfloor+1}^{k\lfloor N/K \rfloor} \bar{g}'_{cN} \quad (2)$$

Where g_{Ak} (1,…,K) represents the gray value of the k^{th} anchor and $\lfloor . \rfloor$ is the floor function. The center pixel and its neighbors are shown in Figure 2 by a yellow rectangle and anchors are computed for K=2. Each of these anchor values are compared with sampling instances of center pixels and binary pattern is generated based on given $s(x)$ condition.These binary strings are further encoded into distinct NLBP codes for different R and P values based on Eq. (3).In this work, R=3 and P=24 are considered and hence 29 NLBP codes are obtained according to U value given by Eq.(4)

$$NLBP_{R,P,K} = \begin{cases} \sum_{p=0}^{P-1} s(\bar{g}_{R,P} - g_{Ak}) \ , U(NLBP) \le 2 \\ P+1 \quad , U(NLBP) = 4 \\ P+2 \quad , U(NLBP) = 6 \\ P+3 \quad , U(NLBP) = 8 \\ P+4 \quad , U(NLBP) = 10 \\ P+5 \quad , Otherwise \end{cases} \tag{3}$$

where R is the local patch radius, P is the number of sampling instances, K is the number of anchors and U is a uniformity measure. U is expressed as

$$U(NLBP) = \left|s(\bar{g}_{R,P-1} - g_{Ak}) - s(\bar{g}_{R,0} - g_{Ak})\right| + \sum_{p=1}^{P-1} \left|s(\bar{g}_{R,P} - g_{Ak}) - s(\bar{g}_{R,p-1} - g_{Ak})\right|$$

$$\text{where } s(x) = \begin{cases} 1, x \ge 0 \\ 0, x < 0 \end{cases}, \ x = \sum_{p=0}^{p-1} (\bar{g}_{R,P} - g_{Ak}) \tag{4}$$

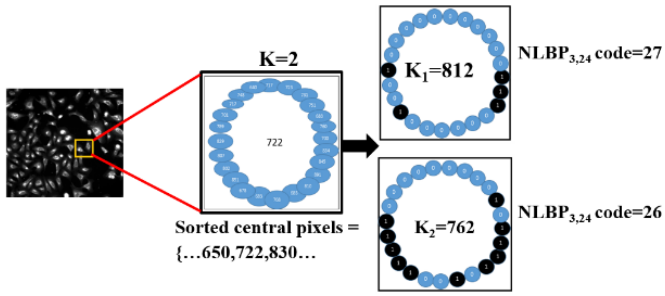


Figure 2. Demonstration of NLBP code generation

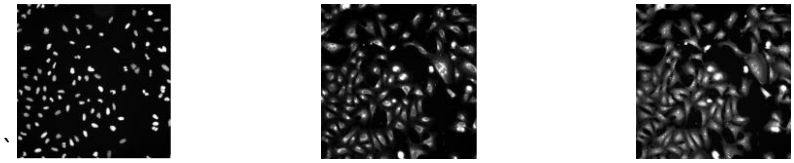
Statistical one-way analysis of variance (ANOVA) is carried out on these NLBP codes to select significant patterns and are fed to the Random forest classifier.

2.3 Classification

Random Forest algorithm is a widely used artificial intelligence technique in medical data classification due to their feature ranking and selection methodology on a random split basis. It splits the feature vectors into different sample sets and builds multiple decision trees with randomly selected features. Each decision tree produces result and finally, the average of multiple tree predictions is taken for decision making [7].

3. Results and Discussion

The representative images of nuclei, endoplasmic reticulum and cytoplasm are shown in Figure 3. It is observed that subtle differences between these microscopic images cannot

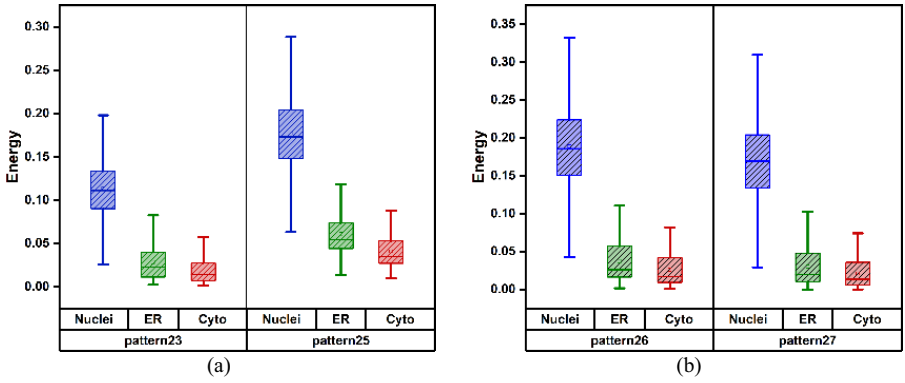


(a) (b) (c)

Figure 3. Representative input images (a) Nuclei, (b) Endoplasmic Reticulum, (c) Cytoplasm

be accessed on visual examination. The box plots shown in Figure 4 (a) and (b) are the energy distribution of representative patterns. It is evident from the box plots of patterns 23,25,26 and 27 that there exists a wide difference between the interquartile range and median values of given organelles. Therefore, the spatial heterogeneity between nuclei, ER and cytoplasm are uncovered with these texture measures.

The extracted NLBP features are analyzed for statistical significance using ANOVA. The patterns 1, 22,23 and 25 to 29 show higher F value which indicates strong inter class variability among cell organelles. It is observed that except pattern 1, majority of significant ($p < 0.05$) patterns capable of discriminating nuclei, ER and cytoplasm have uniformity measure greater than 2. This shows that NLBP quantifies non uniform patterns which corresponds to complex textures such as high curvature edges, lines and corners.



(a) (b)

Figure 4. (a) Box plots of patterns 23 & 25 ,(b) Box plots of patterns 26 & 27

The effect of R and K on classification accuracy are shown in Figure 5(a) and (b). This method achieved an accuracy of 84% with R=3 and K=5.With increase in number of anchors K, NLBP progressively captures intensity variations of local patch with respect to the whole image. The increase in accuracy with increase in R and K is associated with the ability of NLBP to comprehensively capture the underlying spatial relationship in the microscopic images.

Pawlowski N et al. [8] had proposed pre trained neural networks such as ResNet-152, VGG 16, Inception-v3 for extracting meaningful features from cell painted images for classification and obtained an accuracy of 55.34%,66.02% and 70.87% respectively. Goldsborough P et al. [9] had presented Least Squares Generative Adversarial Network (LSGAN) for classification of mechanism of action of chemicals and achieved 68% accuracy. It is observed that the proposed method shows better performance with an accuracy of 84% than the state-of-the art methods. This indicates the clinical relevance of this study in Artificial Intelligence (AI) based solutions for problems in health informatics.

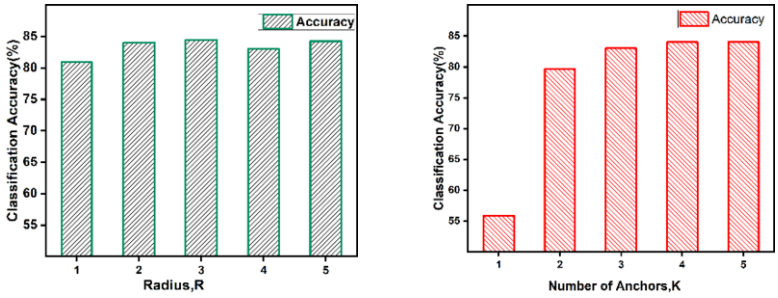


Figure 5. (a) Effect of local patch radius on the classification accuracy, **(**b)Effect of number of Anchors on the classification accuracy

4. Conclusions

In this study, differentiation of cell painted organelles using a Non Local Texture Descriptor with Random Forest classifier is performed. The results demonstrate that NLBP captures non uniform patterns that occur due to wide range pixel interaction in large neighborhoods. This method illustrates the feasibility of global image statistics for effective categorization of cell components in microscopic images. Proper selection of local patch radius and number of anchors are important for accurate classification. This work could be useful for analyzing the effect of cytotoxicity and understanding the reaction of chemical compounds in cell lines. The proposed machine learning based study aids in analyzing the cytological effects with minimum medical expertise.

References

[1] Bray MA, Singh S, Han H, Davis CT, Borgeson B, Hartland C, Kost-Alimova M, Gustafsdottir SM, Gibson CC, Carpenter AE. Cell Painting, a high-content image-based assay for morphological profiling using multiplexed fluorescent dyes. Nature protocols. 2016 Sep;11(9):1757-74.

[2] Fekri-Ershad S. Cell phenotype classification using multi threshold uniform local ternary patterns in fluorescence microscope images. Multimedia Tools and Applications. 2021 Mar;80(8):12103-16.

[3] Yin H, Chen Y, Xiong J, Xia R, Xie J, Yang K. An improved local binary pattern method for pollen image classification and recognition. Computers & Electrical Engineering. 2021 Mar 1;90:106983.

[4] Liu L, Chen J, Fieguth P, Zhao G, Chellappa R, Pietikäinen M. From BoW to CNN: Two decades of texture representation for texture classification. International Journal of Computer Vision. 2019 Jan;127(1):74-109.

[5] Song T, Feng J, Luo L, Gao C, Li H. Robust texture description using local grouped order pattern and non-local binary pattern. IEEE Transactions on Circuits and Systems for Video Technology. 2020 Feb 6;31(1):189-202

[6] Singh S, Wu X, Ljosa V, Bray MA, Piccioni F, Root DE, Doench JG, Boehm JS, Carpenter AE. Morphological profiles of RNAi-induced gene knockdown are highly reproducible but dominated by seed effects. PloS one. 2015 Jul 21;10(7): e0131370

[7] Alam MZ, Rahman MS, Rahman MS. A Random Forest based predictor for medical data classification using feature ranking. Informatics in Medicine Unlocked. 2019 Jan 1;15:100180.

[8] Pawlowski N, Caicedo JC, Singh S, Carpenter AE, Storkey A. Automating morphological profiling with generic deep convolutional networks. BioRxiv. 2016 Jan 1:085118.

[9] Goldsborough P, Pawlowski N, Caicedo JC, Singh S, Carpenter AE. CytoGAN: generative modeling of cell images. BioRxiv. 2017 Jan 1:227645.