Evaluation of Methods for Cell Nuclear Structure Analysis from Microscopy Data

Alexandr A. Kalinin[1,2], Brian D. Athey[1,3], and Ivo D. Dinov[1,2,3]

[1] Department of Computational Medicine and Bioinformatics, University of Michigan Medical School, Ann Arbor, MI 48109, USA,
[2] Statistics Online Computational Resource (SOCR), Health Behavior and Biological Sciences, University of Michigan School of Nursing, Ann Arbor, MI 48109, USA
[3] Michigan Institute for Data Science (MIDAS), University of Michigan, Ann Arbor, MI 48109, USA

Abstract. Changes in cell nuclear architecture are regulated by complex biological mechanisms that associated with the altered functional properties of a cell. Quantitative analyses of structural alterations of nuclei and their compartments are important for understanding such mechanisms. In this work we present a comparison of approaches for nuclear structure classification, evaluated on 2D per-channel representations from a large 3D microscopy imaging dataset by maximum intensity projection. Specifically, we compare direct classification of pixel data from either raw intensity images or binary masks that contain only information about morphology of the object, but not texture. We evaluate a number of widely used classification algorithms using 2 different cross-validation schemes to assess batch effects. We compare obtained results with the previously reported baselines and discuss novel findings.

Keywords: cell nucleus morphology, bioimage analysis, image classification

1 Introduction

Cell nuclear architecture is regulated by complex biological mechanisms related to cell differentiation, development, proliferation, and disease [3, 10, 11]. Changes in nuclear architecture are related to altered functional properties such as gene regulation and expression. Conversely, studies in mechanobiology show that external geometric constraints and mechanical forces that deform the cell nucleus affect chromatin dynamics and gene and pathway activation [9]. Quantitative analyses of structural alterations of nuclei and their compartments (for example, nucleoli) also have medical implications, for example, in detection and treatment of pathological conditions such as cancer [11].

The way cell nuclear phenotypes can be distiguished depends on their representation in the image data [7]. Although a few of signal processing and computer vision algorithms have been proposed to analyze cell and nuclear phenotypes using 3D representations [2], the dimensionality and volume of acquired data,

various image acquisition conditions, and great variability of cells in a population present numerous challenges for 3D image analysis methods that should be scalable, robust to noise, and specific enough across cell populations at the same time. 2D image representations are computationally cheaper to operate on and do often carry enough information to achieve a desired level of performance.

In this work we present a comparison of approaches for nuclear structure classification, evaluated on 2D per-channel maximum intensity projections from a large 3D microscopy imaging dataset. Specifically, we compare direct classification of pixel data from either raw intensity images or binary masks, which contain only object morphology information, but not texture. We evaluate a number of widely used classification algorithms using 2 different cross-validation schemes to assess batch effects. We demonstrate near-perfect classification performance using 2D data and compare our results with originally reported baselines [4].

2 Methods

2.1 Dataset description

In this study we use 3D Cell Nuclear Morphology Microscopy Imaging Dataset [4], one of the biggest publicly available datasets for nuclear structure classification. This dataset consists of two collections of 3D volumetric microscopic cell images with corresponding nuclear and nucleolar binary masks. Each collection includes images of cells in two phenotypic states that have been shown to exhibit different nuclear and/or nucleolar structure. Thus, each collection poses a binary classification problem that can be used for the assessment of cell nuclear and nucleolar phenotype analysis methods. Cells in both collections are labeled with 3 different fluorophores: DAPI (4',6-diamidino-2-phenylindole), a common stain for the nuclei, fibrillarin antibody (anti-fibrillarin) and ethidium bromide (EtBr), both used for nucleoli staining. In the dataset original channels are separated and re-sliced into a $1,024 \times 1,024 \times Z$ lattice ($Z = \{30, 50\}$). Every sub-volume is labeled as c0, c1, c2, representing the DAPI, anti-fibrillarin, and EtBr channels, respectively, **Fig. 1**. For every sub-volume, accompanying vendor meta-data are extracted from the original data. Binary masks in each collection are obtained by segmentation of the original data in c0 and c2 channels, see details in [4].

In this work we focus on the first cell collection in the dataset. It includes images of primary human fibroblast cells. In order to introduce nuclear structural changes changes, a part of this collection was subjected to a G0/G1 Serum Starvation Protocol that is used for cell cycle synchronization and has previously been shown to alter nuclear organization, which was reflected in morphology changes, for example nuclear size and shape [8]. As a result, the first collection contains 178 3D volumetric images of cells in the following phenotypic classes: (1) 64 sub-volumes of proliferating fibroblasts (PROLIF), and (2) 112 sub-volumes of the cell cycle synchronized by the serum-starvation protocol cells (SS). These classes serve as two categories in a binary morphology classification setting.

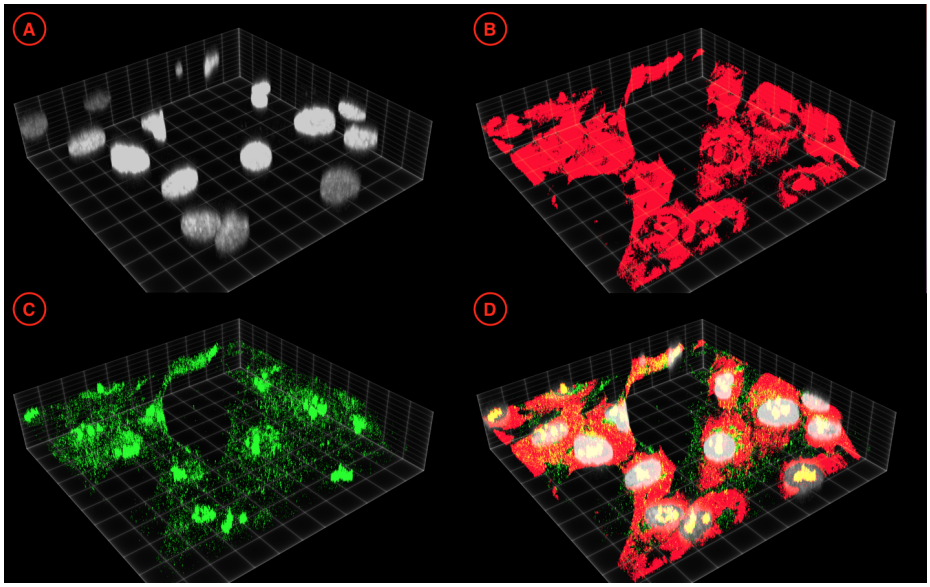

Fig. 1. An exemplar 3D visualization of a data sub-volume from the fibroblast cell collection: (A) DAPI channel; (B) EtBr channel; (C) anti-fibrillarin channel; (D) a composite image. Adapted from [4] under a CC-BY 4.0 International license.

2.2 Data preprocessing

Fluorescent labels are not always specific to the object of interest and often produce noisy background, as shown in **Fig. 1**. In order to accurately assess the changes in the nuclear architecture, we first apply nuclear masks provided with the dataset to all 3 channels of original microscopy imaging data. Due to the anisotropic nature of original data, we then re-scale masked volumes in Z dimension by a factor extracted from the corresponding meta-data. Since each of $1,024 \times 1,024 \times Z$ sub-volumes typically contains between 1 and 5 nuclei, we crop masked and re-scaled volumes into smaller $256 \times 256 \times 57$ sub-volumes, centered at the centroid of the corresponding nuclear mask and zero-pad them, when necessary. Finally, we produce 2D representation of masked, re-scaled, and cropped sub-volumes by calculating maximum intensity projection along the Z dimension. As a result, we create a set of 999 256×256 images per channel. An example of preprocessed images in all 3 channels in shown in **Fig. 2**.

2.3 Classification

We compare various supervised classification algorithms from scikit-learn, a popular Python machine learning toolkit [6], including Gaussian Naive Bayes (NB), Linear Discriminant Analysis (LDA), k nearest neighbors classifier (kNN), support vector machines with linear (SVM) and Gaussian kernels (RBF), Random

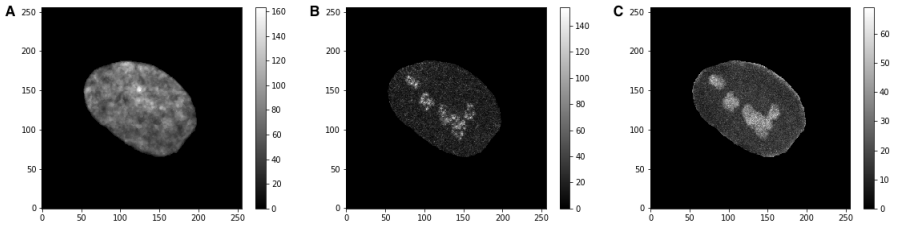

Fig. 2. An exemplar visualization of 256×256 2D maximum intensity projections of a masked, re-scaled, and cropped fibroblast sub-volumes in: (A) DAPI channel, c0; (B) anti-fibrillarin channel, c1; (C) EtBr channel, c2.

Forest (RF), Extremely Randomized Trees (ET), and Gradient Boosting (GBM). All classifiers use default hyper-parameters. Every image is flattened into a 1D feature vector. Feature preprocessing includes subtracting the mean and scaling to unit variance of the training set. We assign the label of the whole image to every single cell extracted from it. In order to assess if batch effects are more apparent in the intensity images compared to binary masks, we compare traditional k-fold cross-validation (CV) scheme with the Leave-2-Opposite-Groups-Out (L2OGO) scheme, suggested in [4]. L2OGO ensures that: (1) all masks derived from one image fall either in the training or testing set, and (2) testing set always contains masks from 2 images of different classes.

3 Results

First, we evaluate the performance of algorithms for fibroblast nuclear classification using only 2D morphological information, i.e. binary masks. We compute AUROC per chennel using 2 different CV schemes: 20 splits in L2OGO and a 10 times repeated 4-fold CV. Results in **Table 1** do not show any apparent batch effects in the 2D classification setting in any of the channels, as performance levels L2OGO are only slightly lower compared to 4-fold CV. As expected, classifiers are not able to pick up complex morphological relationships from flattened binary vectors, even when 3 channels are combined. All such results were dominated by the classification performance using morphometry features manually extracted from binary masks, as described in [4]. The best overall result with L2OGO is achieved by the Gaussian SVM (RBF) classifier in with $AUROC = 0.772 \pm 0.041$, $AUPR = 0.731 \pm 0.063$, and $F1 = 0.682 \pm 0.060$.

Next, we evaluate the performance using only 2D pixel intensity information. Results in **Table 2** demonstrate a possibility of batch effects. The performance on the nuclear c0 channel does not benefit from the presence of additional information compared to only 2D masks. But nucleolar-stained channels c1 and c2 demonstrate 20% gain in performance even using more conservative L2OGO CV. However, L2OGO here leads to a large variance of the performance metric. On average, the EtBr channel (c2) seems to provide a sightly better representation

Table 1. Classification AUROC (*mean* ± *std*) on binary masks for 2 cross-validation schemes (CV: 4-fold and L2OGO) and a number of algorithms (Clf: NB, LDA, kNN, SVM, RBF, RF, ET, GBM) per image channel (c0, c1, c2, and all 3 channels combined). For the reference, the last column contains performance measures obtained using voxel-based morphometry features extracted from binary masks [4].

		Binary images				Morphology [4]
CV	Clf	c0	c1	c2	c0c1c2	c0c2
4-fold	kNN	0.604 ± 0.056	0.615 ± 0.059	0.610 ± 0.059	0.620 ± 0.062	0.706 ± 0.030
	SVM	0.597 ± 0.067	0.583 ± 0.075	0.582 ± 0.054	0.650 ± 0.071	0.646 ± 0.090
	RBF	0.647 ± 0.053	0.648 ± 0.078	$\mathbf{0.644 \pm 0.056}$	0.678 ± 0.062	$\mathbf{0.785 \pm 0.027}$
	RF	0.635 ± 0.061	0.642 ± 0.061	0.635 ± 0.054	0.661 ± 0.064	0.735 ± 0.029
	ET	0.581 ± 0.059	0.577 ± 0.077	0.573 ± 0.058	0.651 ± 0.054	0.732 ± 0.034
	GBM	$\mathbf{0.659 \pm 0.049}$	$\mathbf{0.658 \pm 0.062}$	0.641 ± 0.061	$\mathbf{0.721 \pm 0.054}$	0.783 ± 0.032
L2OGO	kNN	0.582 ± 0.047	0.582 ± 0.047	0.582 ± 0.047	0.582 ± 0.047	0.695 ± 0.030
	SVM	0.557 ± 0.057	0.568 ± 0.050	0.555 ± 0.046	0.564 ± 0.053	0.650 ± 0.077
	RBF	0.616 ± 0.060	0.616 ± 0.060	0.616 ± 0.060	0.616 ± 0.060	$\mathbf{0.772 \pm 0.041}$
	RF	0.635 ± 0.038	0.636 ± 0.037	0.626 ± 0.050	0.627 ± 0.041	0.719 ± 0.039
	ET	0.637 ± 0.039	0.637 ± 0.039	$\mathbf{0.641 \pm 0.044}$	0.636 ± 0.042	0.717 ± 0.034
	GBM	$\mathbf{0.642 \pm 0.037}$	$\mathbf{0.644 \pm 0.038}$	0.639 ± 0.037	$\mathbf{0.643 \pm 0.041}$	0.758 ± 0.041

Table 2. Classification AUROC (*mean* ± *std*) on raw intensity images for 2 cross-validation schemes (CV: 4-fold and L2OGO) and a number of algorithms (Clf: NB, LDA, kNN, SVM, RBF, RF, ET, GBM) per image channel (c0, c1, c2, and all 3 channels combined).

		Raw intensity images			
CV	Clf	c0	c1	c2	c0c1c2
4-fold	kNN	0.581 ± 0.059	0.771 ± 0.048	0.862 ± 0.041	0.865 ± 0.039
	SVM	0.610 ± 0.077	0.726 ± 0.080	0.829 ± 0.059	0.896 ± 0.043
	RBF	0.647 ± 0.058	0.814 ± 0.052	0.892 ± 0.0326	0.938 ± 0.026
	RF	0.630 ± 0.040	0.868 ± 0.039	0.890 ± 0.035	0.948 ± 0.022
	ET	0.606 ± 0.054	0.864 ± 0.045	0.875 ± 0.035	0.961 ± 0.021
	GBM	$\mathbf{0.673 \pm 0.046}$	$\mathbf{0.919 \pm 0.031}$	$\mathbf{0.912 \pm 0.026}$	$\mathbf{0.974 \pm 0.011}$
L2OGO	kNN	0.552 ± 0.030	0.755 ± 0.207	0.826 ± 0.170	0.933 ± 0.044
	SVM	0.579 ± 0.053	0.671 ± 0.166	0.794 ± 0.183	0.964 ± 0.068
	RBF	0.579 ± 0.053	0.766 ± 0.261	0.844 ± 0.204	$\mathbf{0.990 \pm 0.021}$
	RF	0.579 ± 0.053	0.823 ± 0.202	0.841 ± 0.188	0.966 ± 0.057
	ET	0.613 ± 0.047	0.816 ± 0.209	0.839 ± 0.181	0.975 ± 0.034
	GBM	$\mathbf{0.637 \pm 0.045}$	$\mathbf{0.844 \pm 0.235}$	$\mathbf{0.857 \pm 0.204}$	0.990 ± 0.029

of nucleolar structure comared to the anti-fibrillarin (c1). Almost all classifiers in both channels show results superior of those obtained with morphometric features, see **Table 1**. Combining all 3 channels gives the best result, demonstrating the complement nature of stains. The best overall result is achieved by the the Gaussian SVM (RBF) classifier with $AUROC = 0.990 \pm 0.029$, $AUPR = 0.980 \pm 0.040$, and $F1 = 0.877 \pm 0.177$).

4 Discussion

In order to establish baseline evaluation of simple pixel-based nuclear structure classification methods, we provide a comparison of a number of widely used machine learning algorithms on both binary and intensity 2D projections of 3D microscopic images. Although nuclear structure classification performance did not benefit from using texture information, our results indicate usefulness of intensities of nucleolar labels: anti-fibrillarin and EtBr. For nuclei specifically, we recommend using morphometry features extracted from binary mask, since the size and the shape seem to reflect relevant changes. Increased potential for batch effects is only observed in classification of nucleolar structures in channels c1 and c2 separately. Interestingly, combining all 3 channels together seem to alleviate this issue and lead to near-perfect performance in L2OGO scheme.

Presented evaluation has a number of drawbacks and requires further investigation. First, we only use flattened vectors of either binary or raw intensity pixels, while there exist multiple methods for texture feature extraction, which may speed up the calculation. Alternatively, deep learning-based methods can be used for automatic feature learning [1, 5]. Second, we only evaluate performance on 2D maximum intensity projections of 3D images. Bigger study could further address similar issues in the native 3D space. Finally, we assume each nucleus in the same image to be representative of the phenotypic label that is provided for the whole image. This can be addressed by using methods that are robust to label noise.

References

1. Ching, T., Himmelstein, D.S., Beaulieu-Jones, B.K., Kalinin, A.A., Do, B.T., Way, G.P., Ferrero, E., Agapow, P.M., Zietz, M., Hoffman, M.M., Xie, W., Rosen, G.L., Lengerich, B.J., Israeli, J., Lanchantin, J., Woloszynek, S., Carpenter, A.E., Shrikumar, A., Xu, J., Cofer, E.M., Lavender, C.A., Turaga, S.C., Alexandari, A.M., Lu, Z., Harris, D.J., DeCaprio, D., Qi, Y., Kundaje, A., Peng, Y., Wiley, L.K., Segler, M.H.S., Boca, S.M., Swamidass, S.J., Huang, A., Gitter, A., Greene, C.S.: Opportunities and obstacles for deep learning in biology and medicine. Journal of The Royal Society Interface 15(141) (2018)
2. Dufour, A.C., Liu, T.Y., Ducroz, C., Tournemenne, R., Cummings, B., Thibeaux, R., Guillen, N., Hero, A.O., Olivo-Marin, J.C.: Signal processing challenges in quantitative 3-d cell morphology: More than meets the eye. IEEE Signal Processing Magazine 32(1), 30–40 (2015)
3. Jevtić, P., Edens, L.J., Vuković, L.D., Levy, D.L.: Sizing and shaping the nucleus: mechanisms and significance. Current opinion in cell biology 28, 16–27 (2014)
4. Kalinin, A.A., Allyn-Feuer, A., Ade, A., Fon, G.V., Meixner, W., Dilworth, D., de Wet, J.R., Higgins, G.A., Zheng, G., Creekmore, A., Wiley, J.W., Verdone, J.E., Veltri, R.W., Pienta, K.J., Coffey, D.S., Athey, B.D., Dinov, I.D.: 3d cell nuclear morphology: Microscopy imaging dataset and voxel-based morphometry classification results. bioRxiv 208207 (2018), https://doi.org/10.1101/208207
5. Kalinin, A.A., Higgins, G.A., Reamaroon, N., Soroushmehr, S., Allyn-Feuer, A., Dinov, I.D., Najarian, K., Athey, B.D.: Deep learning in pharmacoge-

nomics: From gene regulation to patient stratification. Pharmacogenomics (2018), https://doi.org/10.2217/pgs-2018-0008
6. Pedregosa, F., Varoquaux, G., Gramfort, A., Michel, V., Thirion, B., Grisel, O., Blondel, M., Prettenhofer, P., Weiss, R., Dubourg, V., Vanderplas, J., Passos, A., Cournapeau, D., Brucher, M., Perrot, M., Duchesnay, E.: Scikit-learn: Machine learning in python. Journal of Machine Learning Research 12, 2825–2830 (Nov 2011)
7. Pincus, Z., Theriot, J.A.: Comparison of quantitative methods for cell-shape analysis. Journal of Microscopy 227(2), 140–156 (2007), http://dx.doi.org/10.1111/j.1365-2818.2007.01799.x
8. Seaman, L., Meixner, W., Snyder, J., Rajapakse, I.: Periodicity of nuclear morphology in human fibroblasts. Nucleus 6(5), 408–416 (2015)
9. Uhler, C., Shivashankar, G.: Regulation of genome organization and gene expression by nuclear mechanotransduction. Nature Reviews Molecular Cell Biology 18(12), 717 (2017)
10. Zheng, G., Kalinin, A.A., Dinov, I.D., Meixner, W., Zhu, S., Wiley, J.W.: Rotational 3d mechanogenomic turing patterns of human colon caco-2 cells during differentiation. bioRxiv 272096 (2018), https://doi.org/10.1101/272096
11. Zink, D., Fischer, A.H., Nickerson, J.A.: Nuclear structure in cancer cells. Nature reviews cancer 4(9) (2004)