

Use of cytomorphometry for classification of subcellular patterns in 3D images

Uso da citomorfometria para classificação de padrões subcelulares em imagens 3D

Eduardo H. Silva^{1,2}, Jefferson R. Souza¹, Bruno A. N. Travençolo^{1*}

Abstract: This paper presents a methodology for the classification of subcellular patterns by the extraction of cytomorphometric features in 3D isosurfaces. In order to validate the proposal, we used a database of 3D images of HeLa cells with nine classes. For each cell, several morphological attributes were extracted based on its isosurface. Using the Quadratic Discriminant Analysis (QDA) classifier with the hybrid attribute selector, we achieved 97.59% of accuracy and F1-score of 0.9757 when classifying the subcellular patterns.

Keywords: Image Processing — Cytomorphometry — HeLa cells — QDA

Resumo: Neste trabalho é apresentada uma metodologia para a classificação de padrões subcelulares por meio da extração de características citomorfométricas a partir de isossuperfícies 3D. Para validar a proposta, foi utilizada uma base com imagens 3D de células HeLa contendo nove classes. Para cada célula, vários atributos morfológicos foram extraídos com base na isossuperfície da célula. O classificador Análise Discriminante Quadrática (QDA), juntamente com um seletor de características híbrido, foram utilizados para classificação dos padrões subcelulares, alcançando 97,59% de acurácia e F1-score de 0,9757.

Palavras-Chave: Processamento de Imagens — Citomorfometria — Células HeLa — QDA

¹Faculdade de Computação, Universidade Federal de Uberlândia, Brazil

²Centro Universitário de Patos de Minas (UNIPAM), Brazil

*Corresponding author: travencolo@gmail.com

DOI: <http://dx.doi.org/10.22456/2175-2745.80598> • Received: 26/02/2018 • Accepted: 18/04/2018

CC BY-NC-ND 4.0 - This work is licensed under a Creative Commons Attribution-NonCommercial-NoDerivatives 4.0 International License.

1. Introduction

Proteomics is the study of a set of proteins of a subcellular organelle, tissue or organism. They are responsible for the control of almost all the biological processes [1]. The subcellular localization of proteins is necessary for the construction of models that capture and simulate the cellular behavior, being beneficial for the early diagnosis of diseases.

The localization of a protein may help the scientists to describe their function, for example, a protein located on the plasma membrane may have a transporter role on the ion channel; on the other hand, a cytoplasmic protein may have the catalytic role [2].

In pathology, morphological alterations in cellular structures are essential elements in diagnostic methods. Cytological morphometry or cytomorphometry is a set of morphometry techniques used to describe features of the cellular components [3]. Morphological features of subcellular proteins provide information to recognize subcellular patterns and assist biologists in predicting cellular behavior.

Computational tools have a key role in recent advances in several biological problems [4]. This has led to the emerging area of bioimage informatics, which integrates areas such as Image Processing, Scientific Visualization, Machine learning and others in analysis of biological images [5].

Scientific Visualization consists in applying techniques to transform abstract data from observations into models that can be easily understood for scientific exploration [6]. Marching Cubes is a widely used scientific visualization method in medical imaging for the reconstruction of isosurfaces (i.e., triangular meshes) from 3D volumes [7]. The extracted meshes serve as input in different routines to extract different features from the data. Machine Learning (ML) routines can then be used to identify patterns and perform classification tasks.

In this context, this paper proposes a methodology for extraction of morphological features of subcellular patterns obtained by reconstruction of isosurfaces from 3D images. Machine Learning classifiers were used to learn the subcellular patterns and provide the ability to distinguish them. The obtained accuracy of the proposed work was compared with

previous approaches.

1.1 Related Works

Fluorescence microscopy is a common technique used to detect the spatial distribution of proteins inside a cell. In many cases, the visual inspection of this type of image is not enough to characterize the location of a subcellular pattern. There is a need for a variety of computational methods to perform the localization of subcellular patterns, and several approaches have been proposed in this sense [8].

Automated methods for subcellular location in 2D images from Human Protein Atlas were analyzed in the work by Newberg et al. [9]. The authors obtained an 83% accuracy when considering 45 different tissues. The study proposed by Huh et al. [10] was designed to improve both the computational efficiency and accuracy of the subcellular patterns classification. Gabor filters were used on image patches. The results tested on 20 classes of 2D images classified obtained an overall accuracy of 87.8%, with 2330 images out of 2655 images in the UCSF dataset being correctly classified.

Kheirkhah et al. [11] used a method that improved the features set, which distinguishes subcellular patterns with high accuracy and high speed. This method based on Modified Threshold Adjacency Statistics (MTAS) that is fundamental to threshold the images. The authors used a collection of 2D images of HeLa cells with antibodies and molecular probes against proteins. The Support Vector Machine and 5-fold cross-validation was used and obtained accuracy of 97.06% for MTAS.

The work of Pärnamaa et al. [12] performs the study for automatically detecting the cellular compartment where a fluorescently-tagged protein resides. An 11-layer neural network was used on data from a mapping of yeast proteins, achieving per cell localization classification accuracy of 91%, and per protein accuracy of 99%. The authors identified that the low-level network features correspond to basic image features, while deeper layers separate localization classes. Results demonstrate the usefulness of deep learning for high-throughput microscopy.

Some works in the literature use the HeLa cancer cell database introduced by Velliste et al. [13]. The authors proposed a method for localization of 10 different subcellular patterns in HeLa cells obtained by confocal fluorescence microscopy. Using ML algorithms is possible to make the predictions of the subcellular patterns. In their work, 28 morphological and distance attributes were extracted for validation of the model. The authors used the DNA as a reference in the cell to generate distance-based characteristics, such as the distance from protein center to DNA center. Using Back-Propagation Neural Network (BPNN) classification algorithm, it was obtained 91% of accuracy. In work by Huang et al. [14], using the same 28 morphological attributes, nine were selected by the Stepwise Discriminant Analysis (SDA) method. With the classification algorithm, Majority Voting was obtained 96% of accuracy in their model.

The distance attributes that use DNA as a reference can be problematic because some images may not contain DNA. The images were extracted by different experts in different laboratories, so not always an image contains the DNA [2]. In work proposed by Huang et al. [15], it was used 14 morphological and geometric attributes (the distance features using DNA were eliminated). Using Support Vector Machines (SVM) algorithm for the classification of the samples, the authors obtained 89.1% of accuracy in their proposal.

In work developed by Chen et al. [2], 48 attributes were extracted, consisting of Halarick texture, edge, and morphological attributes. The authors used SDA to select a set of attributes that best discriminate the classes. From the 48 attributes, seven attributes have been chosen, being one edge characteristic, five texture attributes, and the standard deviation of the volume as the only morphological related attribute. With the BPNN algorithm for classification of the samples, the authors obtained 98% of accuracy in their proposal.

The subcellular localization of proteins in 3D images presented promising results over 2D images [14]. One common feature of the methods described in this section is that they have been developed based on the pixel (2D) or voxels (3D), from which several features were extracted, such as, texture, distance, and morphological attributes. In our proposal, on the other hand, we consider the isosurface reconstruction for extraction of attributes and use only the morphological attributes.

2. Materials and Methods

2.1 Dataset

To evaluate our proposal, it was used a database of 3D images of HeLa cells proposed by Velliste et al. [13]. HeLa is cell line derived from a cervical tumor, widely used in several biomedical researches [16]. The database is composed of 454 three-dimensional images (approximately 50 images per class) containing only one cellular component. Each 3D image has about 14 to 24 slices, and the voxel dimensions on axes x , y and z were $0.049 \times 0.049 \times 0.203 \mu\text{m}$ [13]. The images were captured using a confocal laser scanning microscope, focused on locating the following subcellular proteins and organelles: endoplasmic reticulum (ER); two Golgi proteins, giantin (GIA) and gpp130 (GPP); LAMP2 (LAM); mitochondria (MIT); nucleolin (NUC); F-Actin (FAC); tubulin (TUB); and the DNA [13]. The proteins giantin e gpp130 were included in the analysis to evaluate the recall (sensitivity) of the extracted visual features, as these proteins are hard to be distinguished by visual inspection [17].

Figure 1 shows, for each protein and organelle, one image (one slice of the 3D volume). Two channels are presented: the green channel represents the DNA; and the red channel the proteins or organelles. In these images, it is possible to locate subcellular patterns with reference to the DNA.

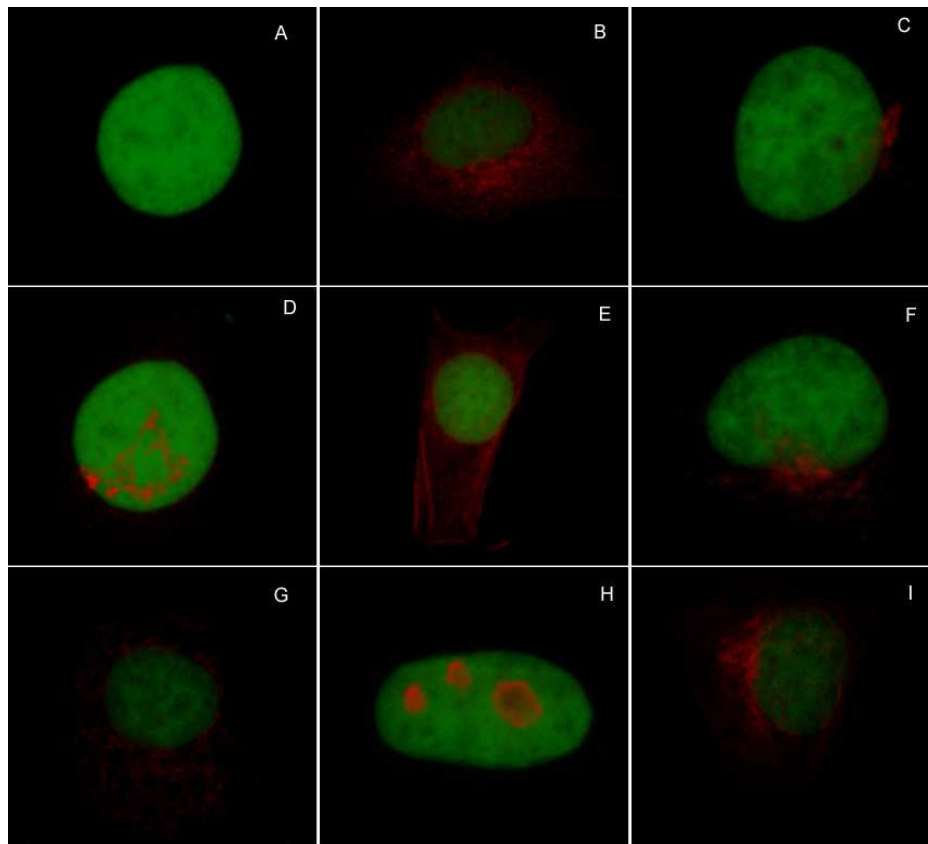


Figure 1. Images samples from the analyzed database. (A) DNA, (B) ER, (C) Giantin, (D) GPP130, (E) F-Actin, (F) LAMP2, (G) Mitochondria, (H) Nucleolin, (I) Tubulin. In all images, the green channel shows the DNA while the red channel shows the proteins or organelles.

2.2 Proposed Methodology

The images were analyzed using the BioWebVis environment [18]. BioWebVis is a web application for cytomorphometric analysis using 3D data. Several algorithms and techniques are available, involving image processing (e.g., pre-processing, segmentation, and feature extraction), computational visualization (e.g., Direct Volume Rendering and Isosurface reconstruction); and machine learning and pattern recognition (e.g., K-Nearest Neighbors, SVM, and Neural Networks) [18].

A general pipeline for the analysis of the data is shown in Figure 2. First, raw data (a set of 2D images) was reconstructed using the Marching Cubes algorithm. This algorithm performs triangulations on volumetric data sets in order to extract surfaces of constant density [7]. After reconstruction, the subcellular proteins and organelles are delimited by an external surface. The extracted surfaces are composed of a polygonal mesh, from which is possible to obtain quantitative information as, for example, the surface area [7]. For all images, the isovalue chosen was 10 (the isovalue is a threshold applied over the voxels intensity to define the location of the surface inside the volume). This threshold value was empirically chosen based on visual analysis.

After that, for each protein or organelle, a total of 17 fea-

tures were computed: surface area, volume, volume on x axis, volume y axis and volume on z axis [19], curvature – mean, standard deviation and entropy [20], Normalized Shape Index (NSI) [21], sphericity, convexity – area and volume [22], coefficients k_x , k_y , and k_z [19] and equivalent spherical diameter [23]. These features were chosen because they describe different morphological aspects of a 3D object and they can be calculated from isosurfaces, avoiding pixel-based measures.

To measure the volume of an object, the use of techniques named Maximum Unit Normal Component (MUNC) and Divergence Theorem Algorithm (DTA) demonstrates superiority over voxel counting [19]. The coefficients k_x , k_y and k_z are the weights factors of the DTA for the volume calculation. These factors are measured as a fraction of the number total of points, where MUNC of the gradients points were directed by the coefficient index, the sum of these three attributes is always equal to 1 [19].

All features computed for each protein or organelle has different scales. A standardization process using Z-score was used in BioWebVis to create a standard scale. Z-Score (z_i) uses the mean (\bar{x}) and standard deviation (s) of the original data (x) to compute the number of variations that the sample is higher or lower than the average value [24], as shown in

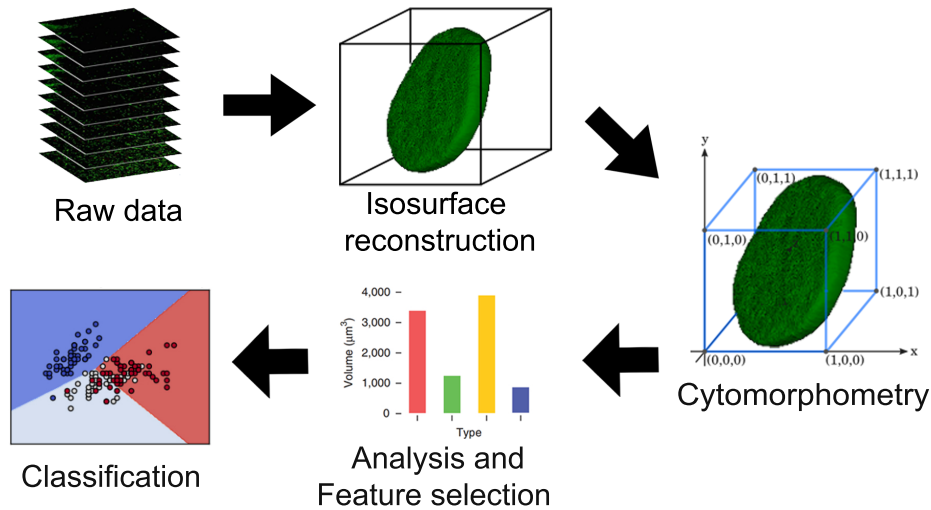


Figure 2. General pipeline used in this work, which can be divided in five main steps: (i) Confocal images are stacked to form the 3D volume (raw images); (ii) Isosurface reconstruction is used to define the polygonal mesh that delimitates the subcellular structure; (iii) Cytomorphometry is performed. In this stage, 17 measures are computed from the isosurface; (iv) Feature selection algorithms are applied in order to reduce dimensionality; (v) Machine learning algorithms are used to classify the subcellular components.

Eq. 1.

$$z_i = \frac{x_i - \bar{x}}{s} \quad (1)$$

After standardization, the data share a common scale, which is essential for pattern recognition. The next step consists in performing feature selection, a process to find the set of features which best separates the classes. In BioWebVis the following techniques are available: variance analysis, regression analysis, Pearson Correlation, Decision Tree [25]. The Pearson coefficient can be computed from the slope of the regression line of two standardized variables – values near 0 indicate low correlation and values near ± 1 mean strong correlation [26].

Next, ML algorithms available on BioWebVis were used to classify the samples and find patterns in the data. K-fold stratified cross-validation was used by dividing the dataset into k mutually exclusive subsets of the same size. One of the subsets is used as testing set, whereas the remaining $k - 1$ subsets are used as training set. This process repeats k times, in which the test subset changes at each turn, to cover all the subsets [27]. In the end, the accuracy and the F1-score are computed. The accuracy expresses the number of correct predictions of the classifier, i.e., the relation of the number of correct predictions and the total number of predictions, as shown in Eq. 2 [28].

$$accuracy = \frac{correct}{predictions} \quad (2)$$

Accuracy is a measure widely used to measure the performance of a classifier. However, if the probabilities of the

classes are too different, it can indicate misleading results. In this case, it is better to use the F1-score, a measure that represents a weighted mean between precision and recall, as expressed in Eq. 3 [28].

$$F1 = 2 * \frac{precision * recall}{precision + recall} \quad (3)$$

Precision measures the potential of the classifier in not predicting as positive the negative objects (Eq. 4).

$$precision = \frac{truepositive}{truepositive + falsepositive} \quad (4)$$

Recall indicates the potential of the classifier in finding all positive samples (Eq. 5).

$$recall = \frac{truepositive}{truepositive + falsenegative} \quad (5)$$

The classification was performed using several types of algorithms: Naive Bayes [29] and the Quadratic Discriminant Analysis (QDA) [30] as classifiers based on Bayes decision theory; SVM [31] as a linear non-probabilistic classifier; K-NN [32] as a classifier based on instances that use distances between samples, and Multilayer Perceptron [33] as a classifier based on Neural Networks.

QDA [30, 34] is a supervised machine learning technique for classification problems, which models the likelihood of each class as a Gaussian distribution, and after that uses the posterior distributions to estimate the class for a given test

Table 1. The result of classifiers using all 17 attributes.

Classifier	F1-score	Accuracy
K-NN	0.9223	92.31%
Multilayer Perceptron	0.9305	93.41%
Naive Bayes	0.9439	94.53%
QDA	0.9391	94.10%
SVM	0.9445	94.51%

point. Furthermore, the Gaussian distribution parameters for each class can be estimated from training data with the maximum likelihood estimation. Lastly, the Gaussian model is satisfied to cases when one does not have much information to characterize a determined class.

3. Results

Figure 3 shows one example of isosurface for each database classes. The green channel represents the DNA and the red channel the proteins or organelles, so it is possible to visualize the location of the subcellular patterns surrounding the DNA. The measurements were extracted from the surfaces of each subcellular pattern and from the DNA.

For analysis, all data were normalized on Z-Score. The results were obtained using the k -fold stratified cross-validation with $k = 10$. It was calculated the mean value of the accuracy and F1-score for each classifier. For K-NN, the first nearest neighbors were used because they presented the best result and in the Multilayer Perceptron classifier 100 neurons in a hidden layer were used. Table 1 shows the result with all 17 attributes extracted using cytomorphometry.

SVM classifier was used considering a linear kernel and $C = 1$ (penalty factor). This classifier obtained the best result, with an accuracy of 94.51% and F1-score of 0.9445. In order to reduce the dimensionality of the attributes and to optimize the classification, we used the decision tree Extra-Trees. The Extra-Trees algorithm [35] is focused on supervised learning problems, which builds an ensemble of regression trees according to the standard top-down procedure. When looking for the best split to separate the samples of a node into two groups, random divisions are drawn for each of the randomly selected features (*maxfeatures*) and best split among those is preferred. When *maxfeatures* is set 1, this amounts to building a random decision tree. Extra-Trees applies several times the learning sample to generate an ensemble model. Lastly, the predictions of the trees are aggregated to create the final prediction, by majority vote in the classification problems. It was selected seven attributes: mean, standard deviation and entropy of the curvature; sphericity; volume in the z axis; and coefficients k_x and k_z . Table 2 shows the results of the classification algorithms with seven attributes.

QDA showed the best result, with an accuracy of 96.50% and F1-score of 0.9649. QDA in combination with the attribute selector improved the accuracy of the classifier over

Table 2. The result of the classifiers using the 7 attributes selected with the Extra-Trees.

Classifier	F1-score	Accuracy
K-NN	0.9231	92.80%
Multilayer Perceptron	0.9318	93.46%
Naive Bayes	0.9419	94.34%
QDA	0.9649	96.50%
SVM	0.9377	94.12%

Table 3. The results of the classifiers using five attributes (mean, standard deviation and entropy of the curvature; volume in the z axis; and coefficient k_z) selected with Extra-Trees and correlation analysis.

Classifier	F1-score	Accuracy
K-NN	0.9368	93.89%
Multilayer Perceptron	0.9319	93.67%
Naive Bayes	0.9460	94.78%
QDA	0.9757	97.59%
SVM	0.9392	93.31%

the previous results (Table 1).

To improve the performance of the proposed model and reduce the dimensionality of attributes, Pearson correlation coefficient was calculated by combining in pairs the seven features. The purpose was to eliminate the linearly correlated attributes. For example, the correlation coefficient between entropy and mean of the curvature is -0.079 and the line of the linear regression graph shown in Figure 4 indicates a low negative correlation between these attributes. Furthermore, the distribution graph in Figure 4 shows that with only these two attributes it is possible to distinguish several samples.

Highly correlated attributes may hinder the classification. For example, the correlation coefficient between curvature entropy and sphericity is -0.957 and the line of the linear regression graph shown in Figure 5 indicates a high negative correlation between these attributes. Also, from the distribution in Figure 5 we noticed the difficulty in distinguishing the classes with only these two attributes.

Based on the correlation between attributes selected by Extra-Trees, five features were selected: mean, standard deviation and entropy of the curvature; volume in the z axis; and coefficient k_z . Table 3 presents the results of the classification algorithms with the five attributes selected. It is worth to note that the correlation analysis alone did not produce satisfactory results when compared to the application of Extra-Trees followed by correlation analysis.

QDA classifier with five attributes selected with a decision tree and statistical analysis presented the best result with accuracy of 97.59% and F1-score of 0.9757. The number of attributes was reduced from 17 to 5, projecting a model of features that better discriminate the subcellular patterns. In

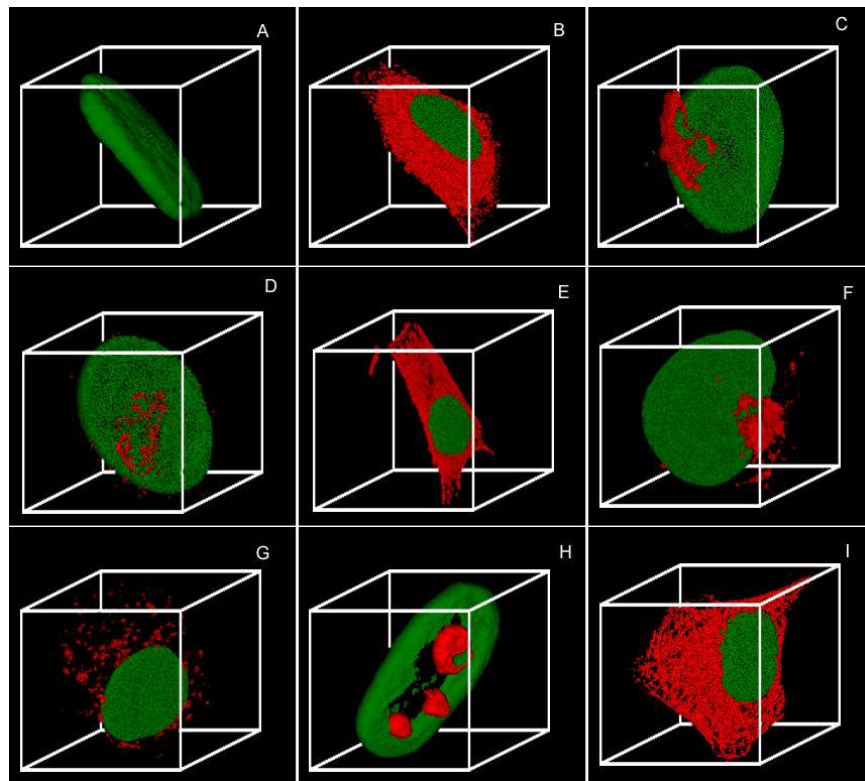


Figure 3. Visualization of the isosurfaces using BioWebVis. The green isosurfaces represent the DNA and the red represent the proteins or organelles. (A) DNA, (B) ER, (C) Giantin, (D) GPP130, (E) F-Actin, (F) LAMP2, (G) Mitochondria, (H) Nucleolin, (I) Tubulin.

this model, we noticed that the curvature features provide a discriminatory capacity relevant to this problem.

Table 4 presents the confusion matrix resulting from the classification with the QDA with five attributes. There is a difficulty in distinguishing the giantin and gpp130 proteins in a visual inspection. The proposed work separated well the two proteins, misclassifying only three samples of these two proteins.

Table 5 shows the result with four subcellular pattern classification approaches using the HeLa 3D data. All approaches have extracted the attributes of voxels from 3D volumes. The works from [13] and [14] obtained the result using the DNA reference for generating of distance-based attributes. In [15], the authors did not use DNA as a reference to create attributes and received a lower result. In [2], a better result was archived when Halarick texture characteristics were included.

Our proposal obtained significant results based on cytomorphometry, without using DNA as a reference to generate attributes. We also included the F1-score to show the weighted average between the precision and recall of the classifier, as a suitable learning algorithm should maximize precision and recall simultaneously. For comparison, we used accuracy (same measure applied in the related works).

4. Conclusions

This paper proposes a methodology for classification of subcellular proteins using cytomorphometric measurements. We present a complete pipeline for analysis 3D images for cell biology, which can be used by cell biology community in their analysis [18]. Due to the recent high availability of 3D images, this type of analysis would become even a more common procedure. The features extraction on 3D isosurfaces may reveal subcellular patterns that are often not seen in a 2D model or in a 3D volume (using voxels instead of surfaces). The curvature features, for example, have been shown to have a high discriminatory capacity of the subcellular patterns, probably due to high variability of the surface shape between the nine analyzed classes.

Our findings reinforce the viability of the construction of biological databases based on polygonal mesh instead of 3D images (voxels), which implies in a considerable reduction in storage and processing. From the polygonal mesh we can extract relevant parameters from the objects. This kind of reduction is extremely important for ML algorithms, reducing the training time. For example, in neural networks the use of pixels/voxels as data input is highly costly.

The reduction of dimensionality allowed the definition of a reduced set of attributes that better discriminate the classes of the model. Using a hybrid selector with a decision tree and statistical analysis, it was possible to achieve the result

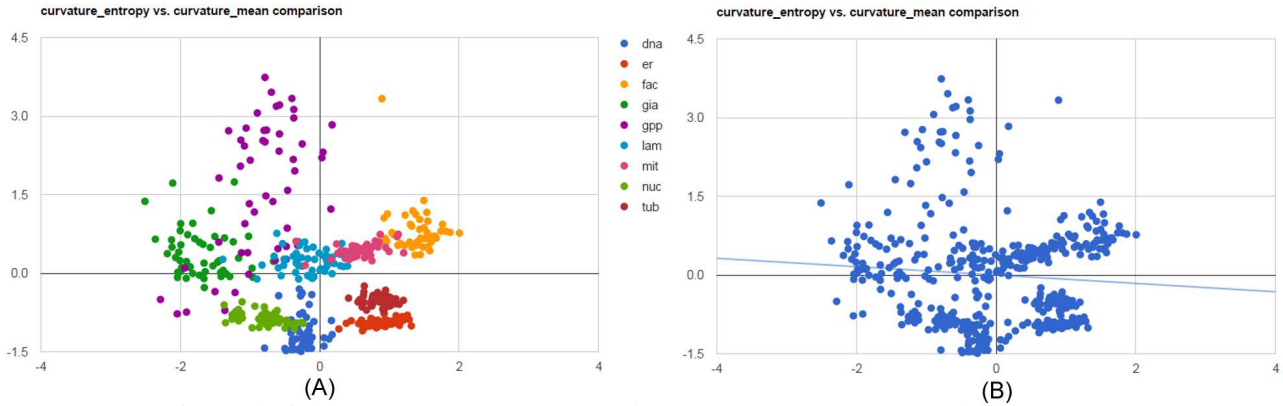


Figure 4. Regression analysis between curvature entropy and mean curvature. (A) Scatter plot between curvature entropy and curvature mean for each class. It is clear from this figure the class separability when considering these two features. (B) Correlation line between the same features, showing a weak correlation (the slope is -0.079).

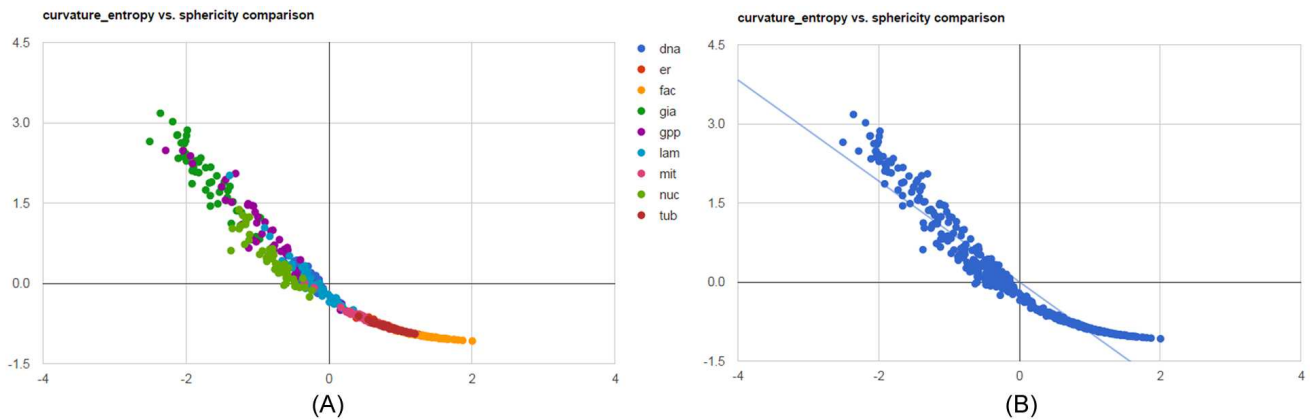


Figure 5. Regression analysis between curvature entropy and sphericity. (A) Scatter plot between curvature entropy and sphericity for each class. The class separation is harder to be identified. (B) Correlation line between the same features, showing a strong negative correlation (the slope is -0.957).

Table 4. Confusion Matrix resulting from the classification using QDA with five attributes.

	DNA	ER	FAC	GIA	GPP	LAM	MIT	NUC	TUB
DNA	52	0	0	0	0	0	0	0	0
ER	1	49	0	0	0	0	0	0	0
FAC	0	0	49	1	0	0	0	0	0
GIA	0	0	0	49	1	0	0	0	0
GPP	0	0	0	2	46	2	0	0	0
LAM	0	0	0	0	3	47	0	0	0
MIT	0	0	0	1	0	0	49	0	0
NUC	0	0	0	0	0	0	0	50	0
TUB	0	0	0	0	0	0	0	0	52

Table 5. Comparison of results with other related works.

Ref.	Classification algorithm	Number of attributes	Number of morphological attributes	Accuracy	F1-score
[13]	BPNN	28	5	91%	
[14]	Majority Voting	9	5	96%	
[15]	SVM	17	5	89%	
[2]	BPNN	7	1	98%	
Our proposal	QDA	5	5	97.59%	0.9757

of 97.59% accuracy and F1-score of 0.9757 with the QDA classifier. The result demonstrates the ability of this methodology to recognize the subcellular patterns and predicting the classification of a new sample. In our proposal is not necessary the prior knowledge of the DNA for the calculation of new attributes. The importance of this finding is that it can reduce the costs of the laboratory analysis, as no DNA staining is necessary. In addition, some databases may not contain the DNA. Therefore, we believe that it can be used in other problems, for example, in analysis of cell deformation in tumors.

The accuracy obtained from our proposal was inferior only to the work by Chen et al. [2], but their work used several types of attributes, not just the morphological ones. For future work, we intend to perform the cytomorphometric analysis in other real databases, as well as to use other measures (visual features) to provide more information to the ML classifiers for learning the cellular patterns successfully.

Acknowledgements

The authors would like to acknowledge the support granted by Fundação de Amparo à Pesquisa do Estado de Minas Gerais – FAPEMIG [grant numbers APQ-01345-13, REONCO-52-11, APQ-03437-15] and Conselho Nacional de Desenvolvimento Científico e Tecnológico - CNPq [grant number 400699/2016-8].

Author contributions

The authors contributed equally to this research.

References

- [1] BARBOSA, E. B. et al. Proteômica: metodologias e aplicações no estudo de doenças humanas. *Rev. Ass. Med. Bras.*, v. 58, n. 3, p. 366–375, 2012.
- [2] CHEN, X.; MURPHY, R. Robust classification of subcellular location patterns in high resolution 3D fluorescence microscope images. In: HUDSON, D.; LIANG, Z.-P. (Ed.). *The 26th Annual International Conference of the IEEE Engineering in Medicine and Biology Society*. Piscataway, NJ: IEEE Service Center, 2004. v. 2.
- [3] WALDEMARIN, K.; BELETTI, M.; COSTA, L. Nuclear morphometry of neoplastic cells as a method for diagnosis of histiocytoma, mastocytoma and transmissible venereal tumor in dogs. *Real Time Imag.*, v. 10, n. 4, p. 197–204, 2004.
- [4] CARDONA, A.; TOMANCAK, P. Current challenges in open-source bioimage informatics. *Nat Methods*, v. 9, n. 7, p. 661–665, 2012.
- [5] MEIJERING, E. et al. Imagining the future of bioimage analysis. *Nat. Biotech.*, v. 34, n. 12, p. 1250–1255, 2016.
- [6] HANSEN, C. et al. *Scientific Visualization*. 1. ed. London, UK: Springer London, 2014. v. 1.
- [7] LORENSEN, W. E.; CLINE, H. E. Marching cubes: A high resolution 3D surface construction algorithm. *ACM SIGGRAPH Comput Graphics*, v. 21, n. 4, p. 163–169, 1987.
- [8] XU, Y.-Y.; YAO, L.-X.; SHEN, H.-B. Bioimage-based protein subcellular location prediction: a comprehensive review. *Front. Comput. Sci.*, v. 12, n. 1, p. 26–39, 2018.
- [9] NEWBERG, J.; MURPHY, R. F. A framework for the automated analysis of subcellular patterns in human protein atlas images. *J. Proteome Res.*, v. 7, n. 6, p. 2300–2308, 2008.
- [10] HUH, S.; LEE, D.; MURPHY, R. F. Efficient framework for automated classification of subcellular patterns in budding yeast. *Cytometry Part A*, v. 75A, n. 11, p. 934–940, 2009.
- [11] KHEIRKHAH, F. M.; HAGHIPOUR, S. Classification of subcellular location patterns in fluorescence microscope images based on modified threshold adjacency statistics. In: *MVIP. 2010 6th Iranian Conference on Machine Vision and Image Processing*. Isfahan, Iran: IEEE, 2010. v. 1.
- [12] PÄRNAMAA, T.; PARTS, L. Accurate classification of protein subcellular localization from high-throughput microscopy images using deep learning. *G3: GENES, GENOMES, GENETICS*, v. 7, n. 5, p. 1385–1392, 2017.
- [13] VELLISTE, M.; MURPHY, R. Automated determination of protein subcellular locations from 3D fluorescence microscope images. In: UNSER, M.; LIANG, Z.-P. (Ed.). *Proceedings IEEE International Symposium on Biomedical Imaging*. Washington, DC, USA: IEEE, 2002. v. 1.

- [14] HUANG, K.; MURPHY, R. F. From quantitative microscopy to automated image understanding. *J Biomed Opt*, v. 9, n. 5, p. 893, 2004.
- [15] HUANG, K.; MURPHY, R. F. Boosting accuracy of automated classification of fluorescence microscope images for location proteomics. *Bioinformatics*, v. 5, n. 1, p. 78, 2004.
- [16] GUILHERMANO, L. et al. *Páginas da história da medicina*. 1. ed. Porto Alegre, RS, Brazil: EDIPUCRS, 2010. v. 1.
- [17] MURPHY, R. F.; VELLISTE, M.; PORRECA, G. Robust numerical features for description and classification of subcellular location patterns in fluorescence microscope images. *J. Vlsi Sig Proc Sysy*, v. 35, n. 3, p. 311–321, 2003.
- [18] SILVA, E. H. et al. Ambiente web para citomorfometria usando imagens 3D. *Health Informatics J*, v. 8, n. 1, p. 643–652, 2016.
- [19] ALYASSIN, A. M. et al. Evaluation of new algorithms for the interactive measurement of surface area and volume. *Med Phys.*, v. 21, n. 6, p. 741–752, 1994.
- [20] MESMOUDI, M. M.; FLORIANI, L. D.; MAGILLO, P. Discrete curvature estimation methods for triangulated surfaces. In: KÖTHE, U.; MONTANVERT, A.; SOILLE, P. (Ed.). *Applications of Discrete Geometry and Mathematical Morphology*. Berlin, Germany: Springer Nature, 2012, (Lecture Notes in Computer Science, v. 7346). p. 28–42.
- [21] SAULTON, A. et al. The role of visual similarity and memory in body model distortions. *Acta Psychol*, v. 164, n. 1, p. 103–111, 2016.
- [22] ZUNIC, J.; ROSIN, P. A new convexity measure for polygons. *IEEE Trans. Pattern Anal. Mach. Intell*, v. 26, n. 7, p. 923–934, 2004.
- [23] KASSIMI, M. A.; ELBEQQALI, O. Semantic based 3D model retrieval. In: ICMCS. *2012 International Conference on Multimedia Computing and Systems*. Tangiers, Morocco: IEEE, 2012. v. 1.
- [24] BACUS, J. W. et al. Image morphometric nuclear grading of intraepithelial neoplastic lesions with applications to cancer chemoprevention trials. *Cancer Epidemiol Biomarkers Prev.*, v. 8, n. 1, p. 1087–1094, 1999.
- [25] LIU, H.; MOTODA, H. *Computational Methods of Feature Selection (Chapman & Hall/Crc Data Mining and Knowledge Discovery Series)*. 1. ed. London, UK: Chapman & Hall/CRC, 2007. v. 1.
- [26] MYERS, J. L.; WELL, A. D.; LORCH Jr, R. F. *Research design and statistical analysis*. 1. ed. London, UK: Routledge, 2013. v. 1.
- [27] PIROOZANIA, M. et al. A comparative study of different machine learning methods on microarray gene expression data. *Genomics*, v. 9, n. Suppl 1, p. S13, 2008.
- [28] JENI, L. A.; COHN, J. F.; TORRE, F. D. L. Facing imbalanced data—recommendations for the use of performance metrics. In: PUN, T.; PELACHAUD, C.; SEBE, N. (Ed.). *Proceedings of the 2013 Humaine Association Conference on Affective Computing and Intelligent Interaction*. Geneva, Switzerland: IEEE Computer Society, 2013. (ACII '13, v. 1), p. 245–251.
- [29] MARTINEZ-ARROYO, M.; SUCAR, L. Learning an optimal naive Bayes classifier. In: TANG, Y. Y. et al. (Ed.). *18th International Conference on Pattern Recognition*. Hong Kong, China: IEEE, 2006. v. 1.
- [30] SRIVASTAVA, S. et al. Bayesian quadratic discriminant analysis. *J. Mach. Learn. Res.*, v. 8, n. 1, p. 1287–1314, 2007.
- [31] HEARST, M. et al. Support vector machines. *IEEE Intell Syst App*, v. 13, n. 4, p. 18–28, 1998.
- [32] DUDANI, S. A. The distance-weighted k-nearest-neighbor rule. *IEEE Trans Syst Man Cybern Syst*, v. 6, n. 4, p. 325–327, 1976.
- [33] THEODORIDIS, S. et al. *Pattern Recognition*. 3. ed. Cambridge, Massachusetts, United States: Elsevier Science, 2006. v. 1.
- [34] HASTIE, T.; TIBSHIRANI, R.; FRIEDMAN, J. *The Elements of Statistical Learning*. 1. ed. New York, NY, USA: Springer New York Inc., 2001. v. 1. (Springer Series in Statistics, v. 1).
- [35] GEURTS, P.; ERNST, D.; WEHENKEL, L. Extremely randomized trees. *Mach. Learn*, v. 63, n. 1, p. 3–42, 2006.