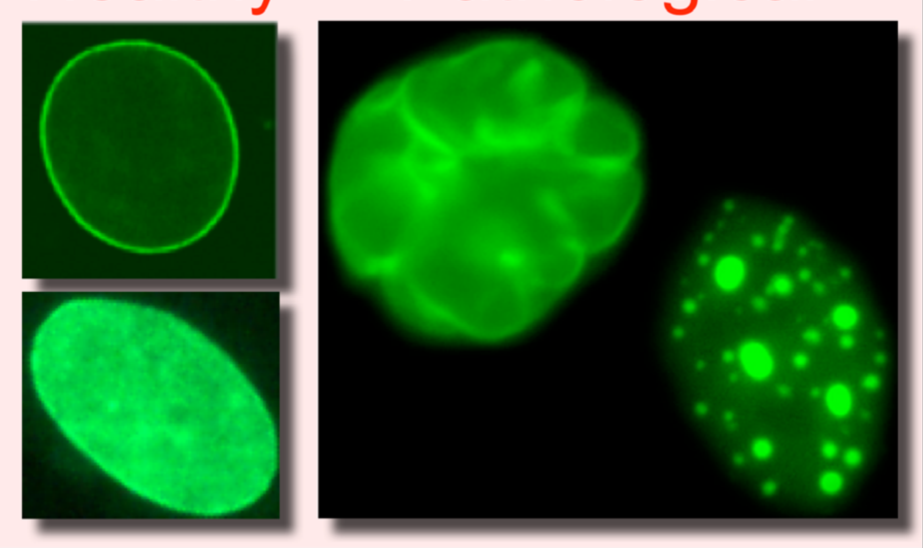
Cells nuclei classification using shape and texture indexes

Guillaume THIBAULT[1], Caroline DEVIC[2], Jean-François HORN[2], Bernard FERTIL[1], Jean SEQUEIRA[1] and Jean-Luc MARI[1]
[1] LSIS laboratory, Images & Models team, Aix-Marseille II University, France
[2] INSERM Unity 678, Paris VI University, France

Motivations: Hudchinson-Gilford syndrome is a rare laminopathy [1] which causes patients age prematurely, failure to thrive and alopecia. Affected patients display a significant proportion of pathological blood cell nuclei: the aim of this poster is to describe a model for nuclei classification. Shape [2] and texture [3] indexes are used in the learning context to build the model.



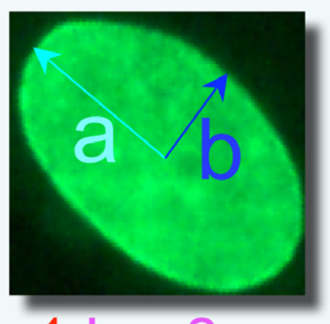
Characterization of the shape

14 classical shape indexes and **Three new shape indexes**

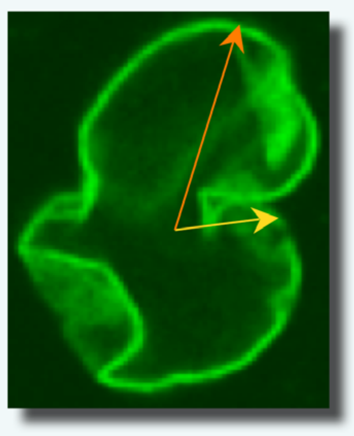
Ellipsoidal shape

Departure from an ellipse

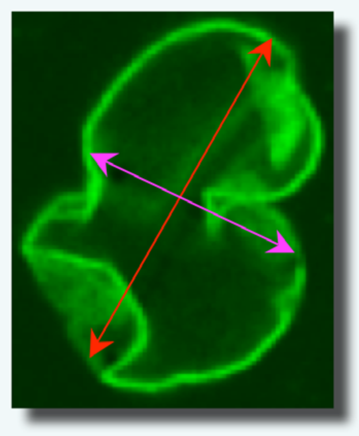
A= $\pi ab \Rightarrow \frac{\pi ab}{A} = 1$



Rmin Rmax



Lap1 Lap2



$$\psi_R = \frac{\pi R_{min} R_{max}}{A}$$

$$\psi_{LAP} = \frac{\pi}{4}\frac{L_{AP1} L_{AP2}}{A}$$

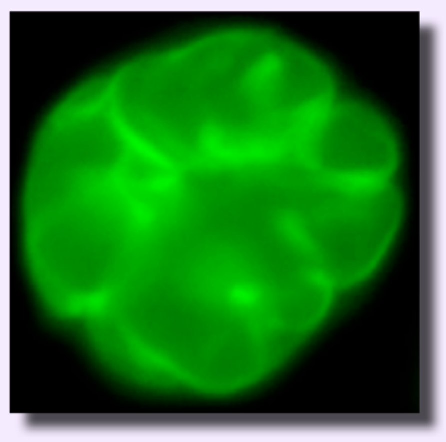
Puffy shape

Gap component scoring

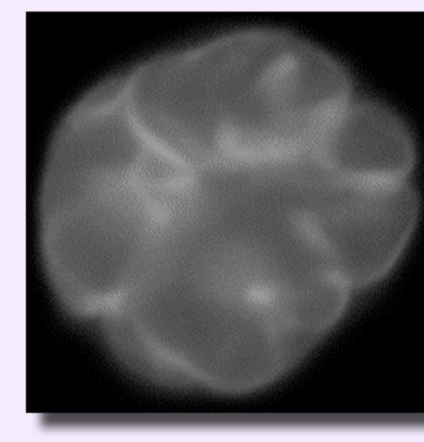
Cce = { ConvexHull(F)\F }

$$\psi_{N_{Cce}} = \frac{1}{1+N_{Cce}}$$

Characterization of the texture



32 grey levels ⇒



⇓

Co-occurrence matrix

$$\begin{bmatrix} m_{1,1} & \cdots & m_{1,32} \\ \vdots & \ddots & \vdots \\ m_{32,1} & \cdots & m_{32,32} \end{bmatrix}$$

Computation of 15 Haralick features ⇐

Validation with the “K-fold” protocol K = 10

A set over 3000 cell nuclei from patients with Hudchinson-Gilford syndrome has been labelled (healthy or pathological, normal shape or puffy, homogeneous or non homogeneous texture) by an expert. It provided the gold standard for learning.

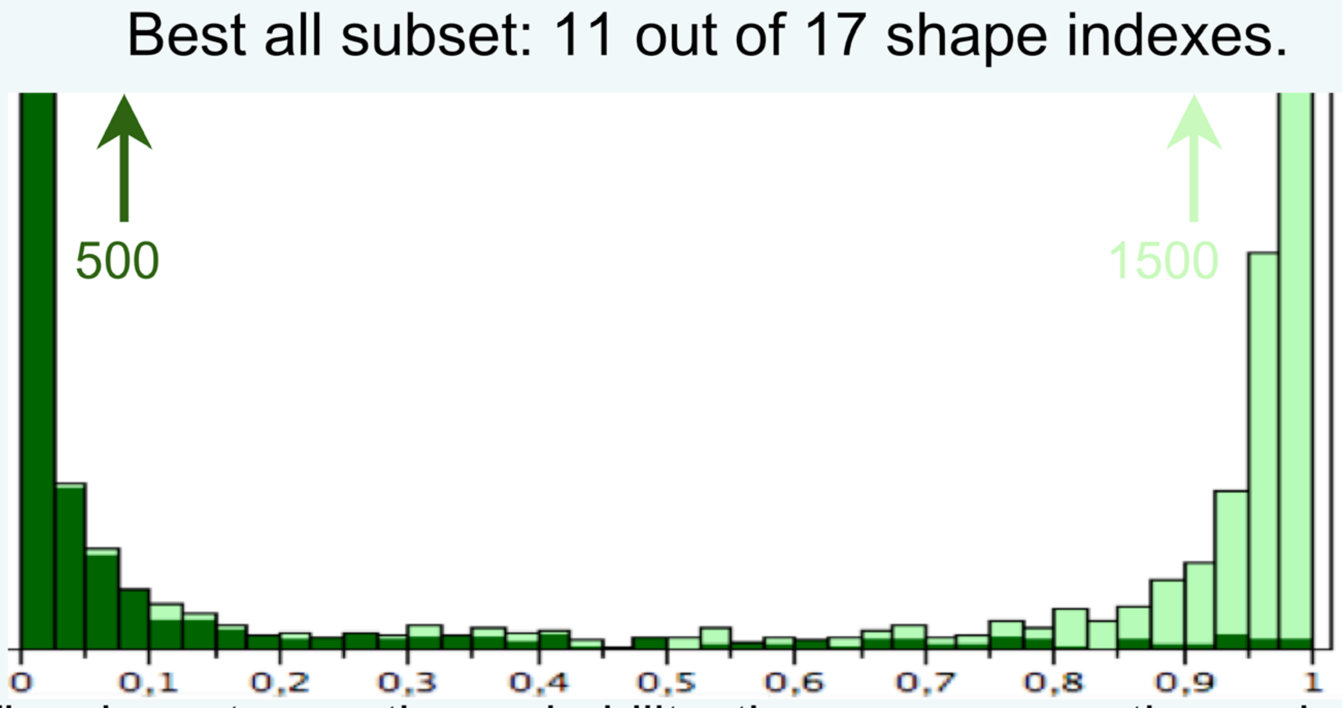
Validation with the “Leave One Out” protocol.

Classification by logistic regression $\frac{e^{f(x)}}{1+e^{f(x)}}$

Distribution of the classification probability as given by the model to the nuclei (validation)

Efficiency of classification of the shape = 95.4%

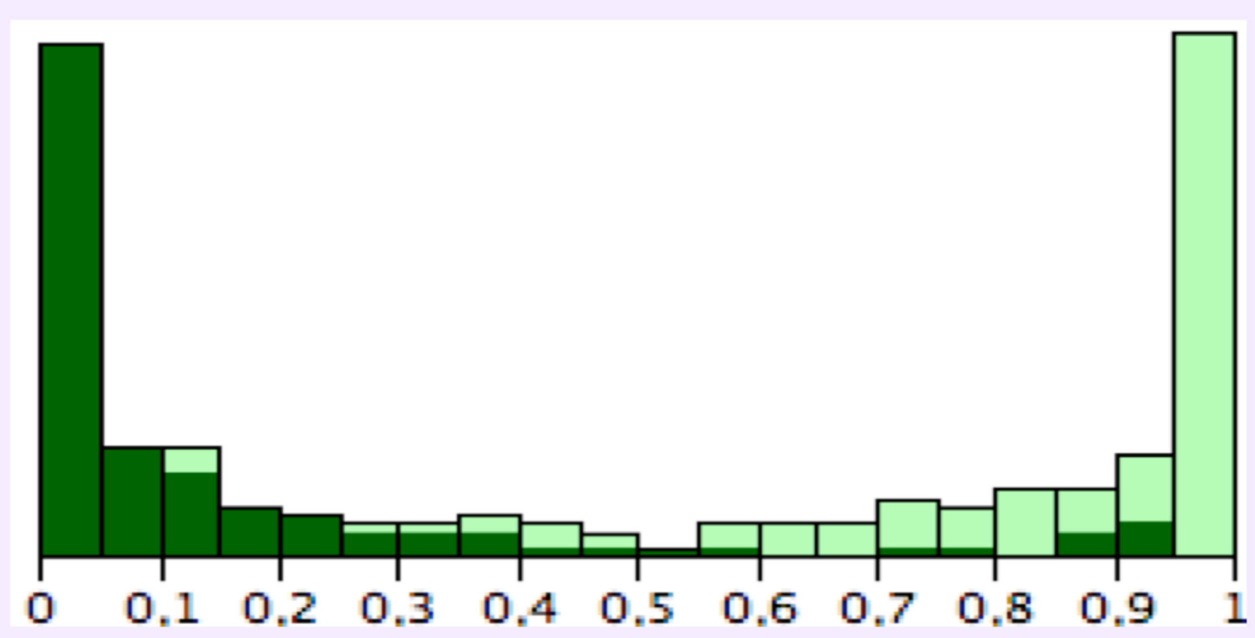
Best all subset: 11 out of 17 shape indexes.



The closer to one the probability, the more convex the nucleus. Dark areas and light green areas are for nuclei labelled as puffy shaped and ellipsoidal shaped respectively.

Efficiency of classification of the texture = 90 %

Best all subset: 8 out of 15 Haralick features.



The closer to one the probability, the more homogeneous the nucleus. Dark areas and light green areas are for nuclei labelled as homogeneous and non homogeneous textured respectively.

Indexes used in the models: shape and textures

Shape indexes

$\psi_{2\ Parallelogram} = \frac{A}{E \times D}$

$Extension_{Diameter} = \frac{E}{D}$

$Extension_{Radius} = \frac{\rho_i}{\rho_e}$

$Circularity = \frac{R_{min}}{R_{max}}$

$Deficit = 1 - \pi\frac{(\rho_e - \rho_i)^2}{P^2}$

$Convexity_{Perimeter} = \frac{P(ConvexHull(F))}{P(F)}$

$Convexity_{Surface} = \frac{A(F)}{A(ConvexHull(F))}$

$Symmetry_{Besicovitch} = \sup_{x \in F} \frac{A(F \cap Symmetric(F,x))}{A(F)}$

Haralick features

standard deviation $\sigma = \sum_x \sum_y (p(x,y) - m)^2$

correlation $\sum_x \sum_y (x-m)(y-m)p(x,y)/\sigma^2$

average of the sums

entropy of the sums

entropy $\sum_x \sum_y p(x,y) log(p(x,y))$

standard deviation of the differences

homogeneity $\sum_x \sum_y \frac{1}{1+|x-y|} p(x,y)$

dissimilarity $\sum_x \sum_y |x-y| p(x,y)$

Conclusions

- We have constructed three new shapes indexes for characterizing the shape of the nuclei. With these indexes, we obtain 95.4% of efficiency of classification and only 93.7% without.
- A model based on Haralick’s features was built in order to handle the problem of texture characterization. This model has provided a satisfactory handling of the texture characterization (90%).
- With the combination of the two models (shape and texture), the efficiency of classification of nuclei as healthy/pathological reaches 90%. It matches the expert's reproducibility.

References

[1] Annachiara De Sandre Giovannoli, Rafaelle Bernard, Pierre Cau, Claire Navarro, Jeanne Amiel, Irene Boccaccio, Stanislas Lyonnet, Colin L. Stewart, Arnold Munnich, Martine Le Merrer, and Nicolas Levy. Lamina truncation in progeria. *Science*, 300(5628):2055, 2003.
[2] Michel Coster and Jean-Louis Chermant. *Précis d’analyse d’images*. Editions du CNRS, 1985.
[3] R. M. Haralick. Statistical and structural approaches to texture. In *Proceedings of the IEEE*, volume 67, pages 786–804, 1979.

February 4-8 2008
WSCG 2008