

# A CAD System for the Acquisition and Classification of Breast TMA in Pathology

M. Milagro Fernández-Carrobles<sup>a,1</sup>, Gloria Bueno<sup>a</sup>, Oscar Déniz<sup>a</sup>, Jesús Salido<sup>a</sup>,  
Marcial García-Rojo<sup>b</sup>, Lucía González-López<sup>c</sup>

<sup>a</sup> *VISILAB, Universidad de Castilla-La Mancha, Spain*

<sup>b</sup> *Department of Pathology, Hospital de Jerez de la Frontera, Spain*

<sup>c</sup> *Department of Pathology, Hospital General Universitario de Ciudad Real, Spain*

**Abstract.** Breast cancer is the most common type of cancer and the fifth leading cause of death in women over 40. Therefore, prompt diagnostic and treatment is essential. In this work a TMA Computer Aided Diagnosis (CAD) system has been implemented to provide support to pathologists in their daily work. For that purpose, the tool covers each and every process from the TMA core image acquisition to their individual classification. The first process includes: tissue core location, segmentation and rigid registration of digital microscopic images acquired at different magnifications (5x, 10x, 20x, 20x and 40x) from different devices. The classification process allows performing the core classification selecting different types of color models, texture descriptors and classifiers. Finally, the cores are classified into three categories: malignant, doubtful and benign.

**Keywords.** High-dimensional image analysis, Tissue microarray (TMA) core, texture descriptor, colour model, classifier.

## Introduction

TMA's are able to concentrate in a paraffin block several hundred cores of tissue in an ordered array. These cores are histological sections used to detect gene expressions or chromosomal alterations which allow carrying out human cancer analysis. However, the use of TMA generates large amounts of information, which requires considerable time. Nowadays, pathologists do this careful work spending significant time under the microscope which can cause fatigue. Besides, the results are subject to subjective interpretations of the pathologist, differing in doubtful or difficult cases. Hence the importance of developing an automatic system which provides support to pathologist in its daily work. Our aim was to develop an automatic tool able to perform TMA core acquisition and classification. To this end, a complete study on breast TMA based on color and texture descriptors was carried out.

---

<sup>1</sup> M. Milagro Fernández Carrobles.

## 1. Methods

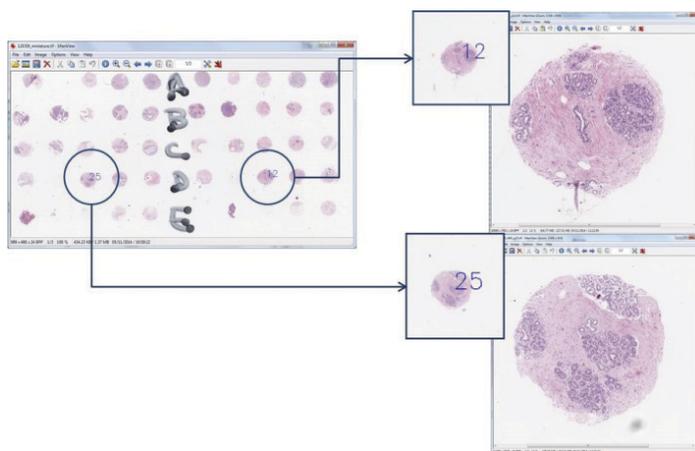
The TMA CAD system is composed of two distinguishable options. The first option allows performing an automatic segmentation, extraction and archiving of the TMA cores at different magnifications and with different devices. Once the TMA cores have been acquired these can be classified using the second tool option. Extraction and classification methods are described in what follows.

### 1.1. TMA core acquisition

The acquisition of the TMA digital images has been carried out using the robotized microscope ALIAS II and the Aperio ScanScope T2 at different magnifications (1.24x (thumbnail), 5x, 10x, 20x and 40x). These images correspond to the whole TMA image stained with hematoxylin and eosin (H&E). The TMA CAD system performs the task of extracting each core individually.

The core acquisition process can be divided into 2 sub-processes: selection and extraction. The selection process performs core segmentation on the TMA. For that purpose, the TMA cores are detected and selected in the thumbnail image. The core segmentation is carried out by different image processing algorithms.

Once the contour cores have been detected they must be analyzed and selected for later extraction. This selection has to be carried out because there are some contours that correspond to noise and imperfections on the glass slide (fragmented, dirt, glue or air bubbles produced in the preparation process or missing core sections). The conditions to consider a given core as valid or not are based on the amount of segmented tissue and contour size. If the percentage of background pixels is lower than 82% the core is considered as valid. This percentage is high since there are cores with low staining, low amount of tissue or contain fat tissue.



**Figure 1.** Individual core extraction.

When the TMA cores have been selected they go through the positioning and extraction processes. Cores are positioned depending on their coordinates in the TMA. They can be easily localized by the pathologist. After finishing the positioning process the TMA cores are extracted. For that, it is necessary to know the relationship between

the TMA thumbnail image and the rest of magnifications. Then, the coordinates of each core are calculated for 5x, 10x, 20x, 40x.

Another problem in the core extraction is the fact that the TMA image is not extracted in only one image. When the microscope extracts the TMA image this is divided into several tiles. For this reason, the cores cannot be located in the same tile and their corresponding tiles must be stitched together [1].

Finally, each segmented core is saved as an individual image. A directory hierarchy is created for each TMA acquired. Inside the parent directory, whose name is the TMA ID, there are five subdirectories: thumbnail, 5x, 10x, 20x and 40x. Each subdirectory contains the TMA core images corresponding to its magnification. The thumbnail subdirectory also contains the thumbnail image with the positioning cores. The size of the TMA core images ranges from about 18 KB for thumbnail cores to 120 MB for 40x.

Besides, a new option was added to the tool since pathologists do not necessarily want to extract all the cores in a TMA. Therefore, the TMA CAD system has two ways to extract the cores, that is, extract all the TMA cores automatically (new TMA cases) or only extract some cores selected by their coordinates and positions, see Fig. 1.

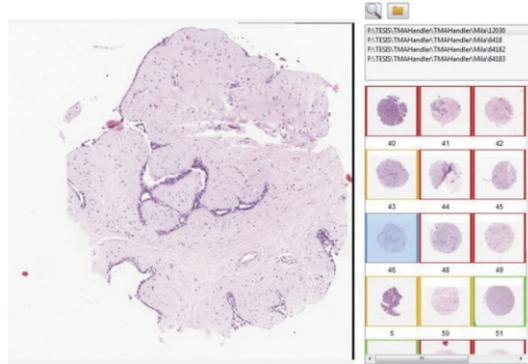
### *1.2. TMA core classification*

The whole TMA core is not used directly in the classification because the classification result would be wrong. A core may contain different types of tissue: stromal, adipose tissue, lobules and in-situ ductal carcinoma. Therefore, the core is previously divided into several regions of 200x200 pixels with an overlap of 100 pixels to make an appropriate selection of the tissue structures. The training set used in the classification is composed of significant interest regions of 4 tissue classes. TMA tissue classes selected were: 1) benign stromal tissue with low and medium cellularity, 2) adipose tissue, 3) benign structures and anomalous and 4) malignity, ductal and lobular carcinomas.

TMA core images are acquired in an RGB color model. In this respect, there are some studies about the influence of the color model in histopathological images [2, 3, 4]. For that reason, The TMA CAD system allows to perform the classification using six different color models: RGB, CMYK, HSV, Lab, Luv, SCT (spherical coordinate transform), two combinations of channels: Lb and Hb and one combination of color models, that is, LuvSCTHb. Once the TMA core regions have been converted to the color model selected they are filtered by a texture descriptor. As in the case of the color models, the TMA CAD system allows choosing between six types of texture descriptors that is, intensity, Fourier, Wavelets, Gabor [5], ILBP (Improved Local Binary Patterns) [6] and frequential [7] and spatial textons [8]. The number of filtered images depends on the descriptor selected. Fourier, Wavelets and Gabor descriptors compute 3 filtered images per original region. After the core regions are filtered, 1st and 2nd order statistical descriptors [9] are extracted from these images. Finally, The TMA CAD system obtains an average of 229 features using intensity, ILBP and textons and an average of 687 features using Wavelets and Gabor. Once feature extraction is finished, the classification is carried out. The classification can be performed using one of the following tree classifiers: gradient boosted trees, random forest or SVM (support vector machine).

### 1.3. The TMA CAD core viewer

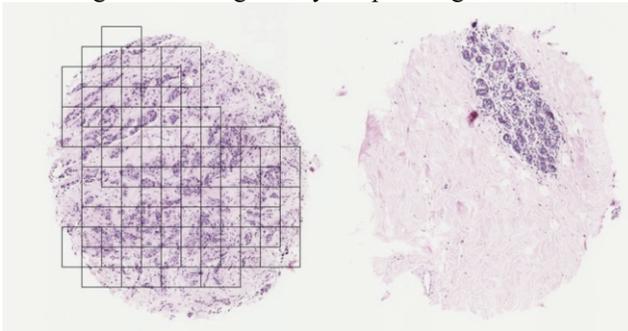
Besides acquiring and classifying TMA cores, the TMA CAD system has a viewer where users can visualize the TMA cores and their classification result. When the user selects a specific TMA, their thumbnail core images are loaded in a table. This table shows the core images with a color bounding box which indicates the classification result: red (malignancy), orange (suspicious) and green (benignant), see Fig. 2. A new window with different options of zoom is opened when the user clicks on the core image.



**Figure 2.** The TMA CAD system core viewer

## 2. Results

This section is focused on the classification results obtained with the TMA CAD system. 228 cores corresponding to 7 TMA samples stained with H&E were classified. These cores were previously classified by a pathologist, being 106 benign, 11 doubtful and 112 malignant. The best classification results were obtained using the RGB color model with intensity and using the gradient boosted trees classifier. The confusion matrix obtained with this classification is shown in Table 1. An average 59.53% precision (positive predictive value) and 89.3% accuracy were obtained. The low precision value is mainly due to the fact that the doubtful cores were not classified as doubtful. However, this is the less important class because these cores have not even been classified as malignant or benignant by the pathologist.



**Figure 3.** Final TMA core image

**Table 1.** Confusion matrix of the best classification

Label	B	D	M	Positive Predictive Value (%)	Negative Predictive Value (%)	Sensitivity (%)	Specificity (%)	Accuracy (%)
B	87	2	2	95.6	86.1	82.07	96.72	89.9
D	7	0	4	0	94	0	94	90
M	12	9	105	83.33	94	94.59	82	88.15

During classification, the TMA CAD system creates an image of each classified core. This image shows the malignant zones inside red rectangles, see Fig. 3. Later, the pathologist may consult these images to know where the conflicting areas are.

### 3. Discussion

This paper shows the result of several years studying histological images of breast TMA. The TMA CAD system is a tool which provides support to pathologists in their daily work. The tool allows automatically extracting the TMA cores at different magnifications and then with the same images at 10x perform tissue classification. These classifications are based in four tissue classes: benign stromal with cellularity, adipose tissue, benign structures and anomalous and ductal and lobular carcinomas. In this way, when the tissue has been classified, the TMA CAD system can decide if the core is malignant, benign or doubtful. For this task, the pathologist can select between nine color model combinations, six types of texture descriptors and three classifiers. The best results were obtained using the RGB color model with intensity images and the gradient boosted trees classifier. These results showed an average 59.53% precision and 89.3% accuracy respectively making the TMA CAD system a useful tool in breast TMA core classification.

### References

- [1] C. Aguilar, M. Fernández, J. Vidal, N. Vález, O. Déniz, J. Salido and G. Bueno. Unión Automática de Imágenes Microscópicas de alta Resolución. Congreso Anual de la Sociedad Española de Ingeniería Biomédica (CASEIB), Proceedings (2011), 10-17.
- [2] M. Mete and U. Topaloglu, Statistical comparison of color model-classifier pairs in hematoxylin and eosin stained histological images, IEEE Symposium on Computational Intelligence in Bioinformatics and Computational Biology, Proceedings (2009), 284–291.
- [3] A. Tabesh, M. Teverovskiy, Tumor classification in histological images if prostate using color texture, ACSSC'06, Fortieth Asilomar Conference on Signals, Systems and Computers, Proceedings (2006), 841-845.
- [4] M. DiFranco, G. OHurleyb, E. Kayc, R. Watsond, P. Cunningham, Ensemble based system for whole-slide prostate cancer probability mapping using color texture features, Computerized Medical Imaging and Graphics, **35** (2011), 629-645.
- [5] J.K. Kamarainen, V. Kyrki, and H. Klviinen, Invariance properties of Gabor filter-based features—overview and applications, IEEE Transactions on Image Processing **15**, no. 5 (2006), 1088–1099.
- [6] H. Jin, Q. Liu, H. Lu, and X. Tong, Face detection using improved LBP under Bayesian framework, Third International Conference on Image and Graphics (ICIG'04), Proceedings (2004), 306–309.
- [7] T. Leung, J. Malik, Representing and recognizing the visual appearance of materials using three-dimensional textons, International Journal of Computer Vision **43**, no. 1 (2001), 29-44.
- [8] M. Varma, A. Zisserman, Texture classification: are filter banks necessary? IEEE Computer Society Conference on Computer Vision and Pattern Recognition **2** (2003), 691-698.
- [9] R. M. Haralick, K. Shanmugam, and I. Dinstein, Textural features for image classification, IEEE Transactions on Systems, Man, and Cybernetics **3**, no. 6 (1973), 610 – 621.