12th European Congress on Digital Pathology
previously European Congress on Telepathology and International Congress on Virtual Microscopy
Paris, France, 18-21 June 2014
Collège des Bernardins, close to île Saint-Louis

Frequential versus Spatial Colour Textons for Breast TMA Classication

Engineering School, University of Castilla-La Mancha

Patology Dpt. Hospital General de Ciudad Real

Gloria Bueno, M Milagro Fernandez-Carrobles,
Oscar Deniz, Jesus Salido,
Marcial García-Rojo

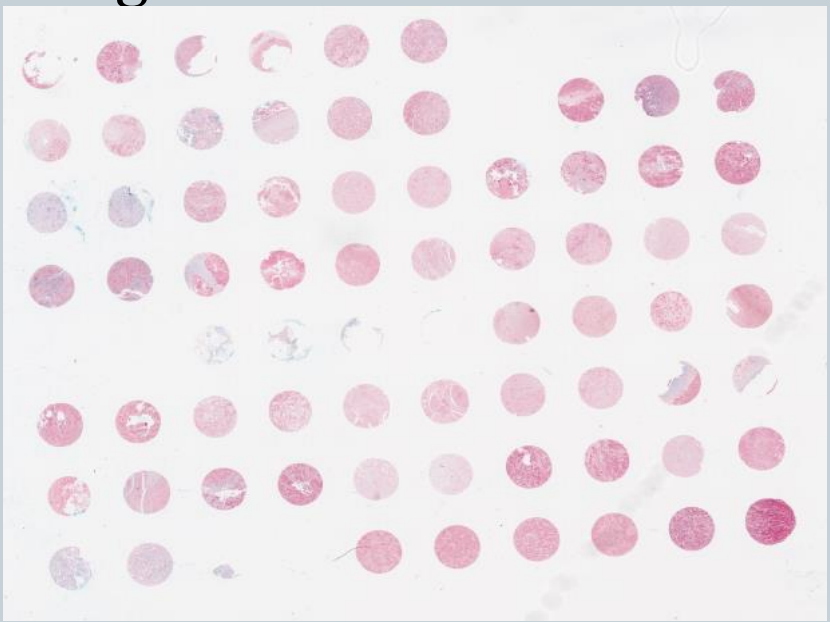
European Commission
RESEARCH & INNOVATION
Marie Curie Actions
AID PATH
UCLM UNIVERSIDAD DE CASTILLA-LA MANCHA
E.T.S. Ingenieros Industriales Ciudad Real
VISILAB GRUPO DE VISIÓN Y SISTEMAS INTELIGENTES

Contents

Background

- The tissue Microarray (TMA) is an ordered array of up to several hundred small cylinders of single tissues (core sections) in paraffin block from which sections can be cut and processed like any other histological section.

- The aim is to automaticaly evaluate the TMA.



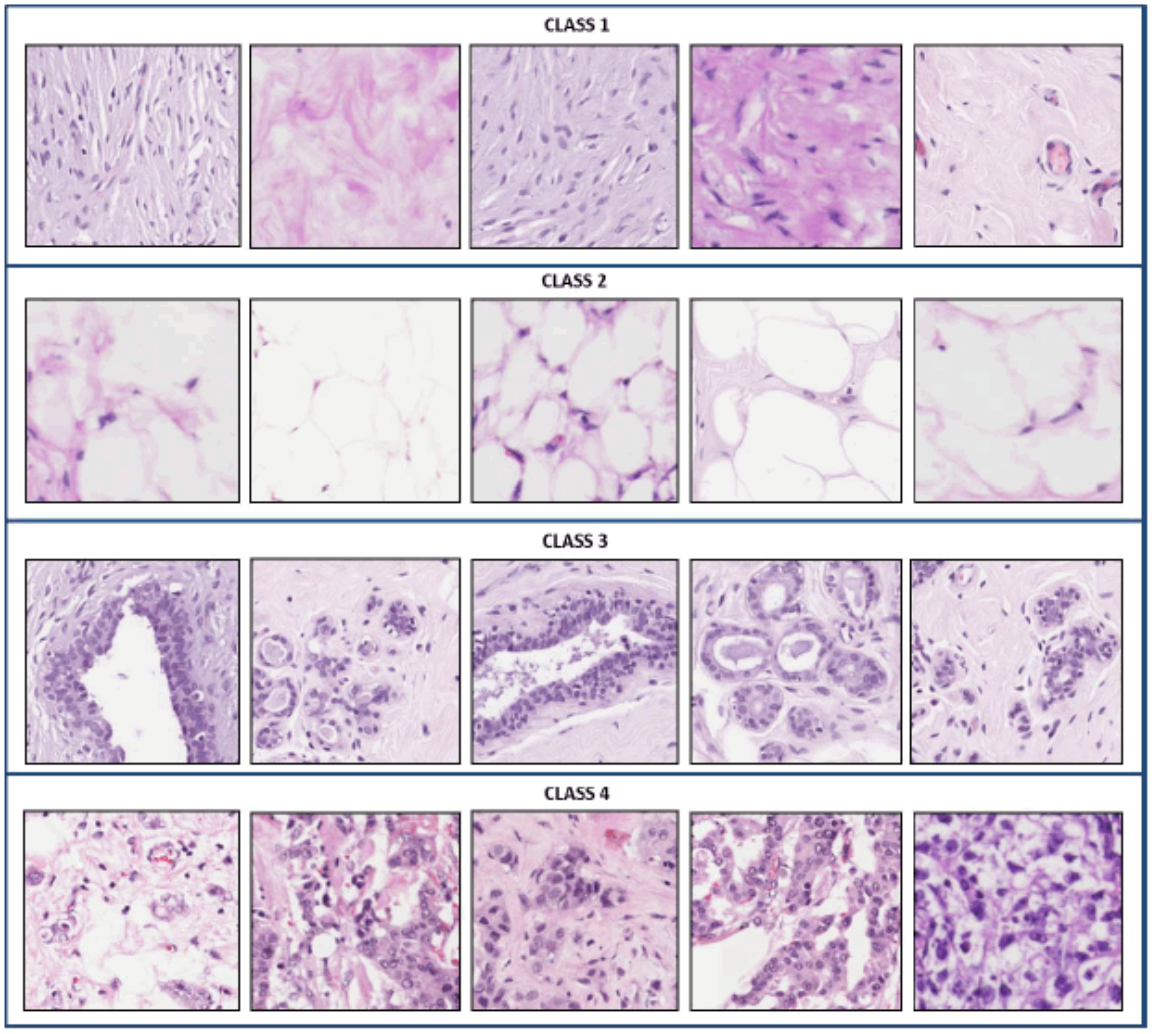
Background

- **Texture** may be defined as the variation of data at scales smaller than the scales of interest. Texture is an important cue in object recognition as it tells us something about the structure from which the objects are made.
- Textons have been selected in our study due to their capability to represent texture on images with by different levels of illumination, distortions or rotations.
- This capability is very important when working with TMA images obtained from a microscope or scanner.

Methods and Materials

- A data set composed by 40 TMA images stained with H&E was digitalized with different devices.
- 628 representative regions (200x200 pixeles) of **4 tissue classes** were selected:
  - 1) benign stromal tissue with low and medium cellularity
  - 2) adipose tissue
  - 3) benign structures but anomalous: sclerosing and adenosis lesions, fibroadenomas, tubular adenomas, phyllodes tumors, columnar cell lesions and duct ectasia
  - 4) different kinds of malignity: ductal and lobular carcinomas.

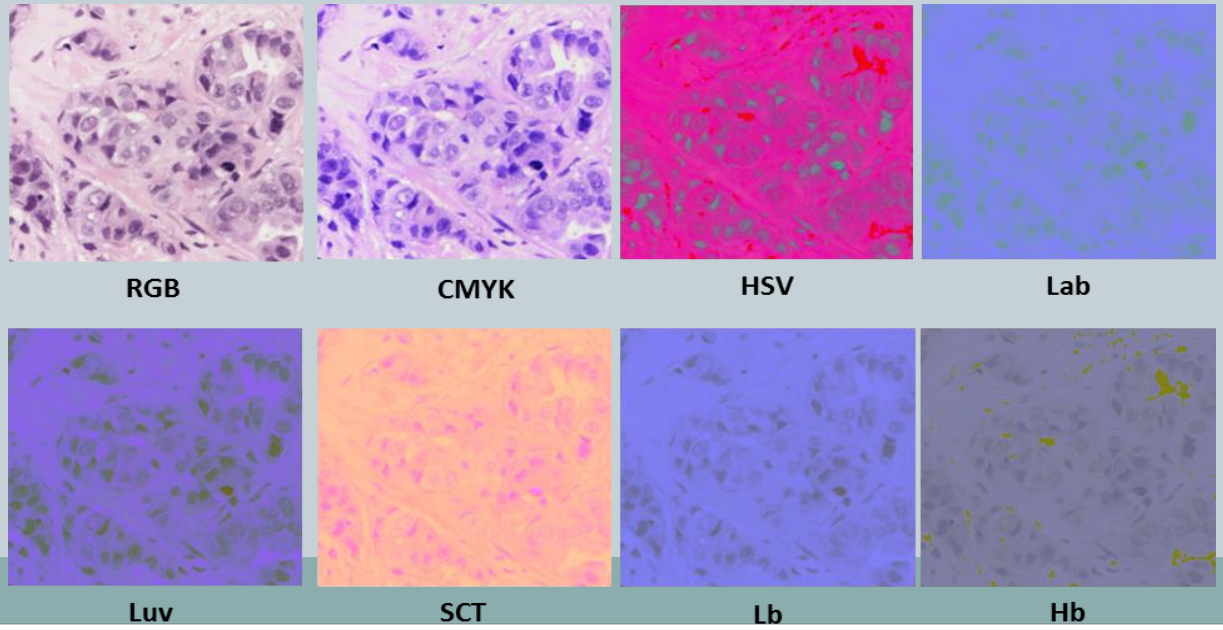
Materials



- 1) benign stromal tissue with cellularity
- 2) adipose tissue
- 3) benign structures but anomalous: sclerosing and adenosis lesions, fibroadenomas, tubular adenomas, phyllodes tumors, columnar cell lesions etc
- 4) malign areas: ductal and lobular carcinomas.

Colour Models

- Colour is not interpreted in the same way by human eyes than by a computer.
- In our study we are used 6 types of colour models, that are, RGB, CMYK, HSV, Lab, Luv, SCT and 2 combinations of them, that are, Lb and Hb.

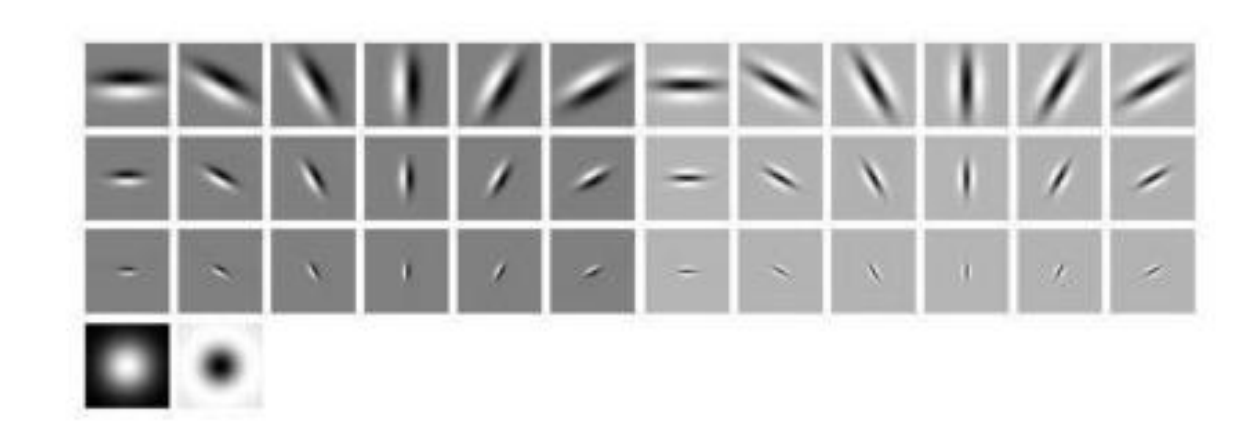


Textons

- Breast tissue contains different types of textures. These textures can be used as discriminators in order to detect cancer.
- Textures are represented by maps which have been generated from the textons, which allow to perform a posteriori classification on the textures analysed.
- In this study two types of textons were used: frequential and spatial textons:
  - Frequential textons use filter banks
  - Spatial textons are computed from an NxN square neighbourhood around each pixel of the original image.

Frequential Textons

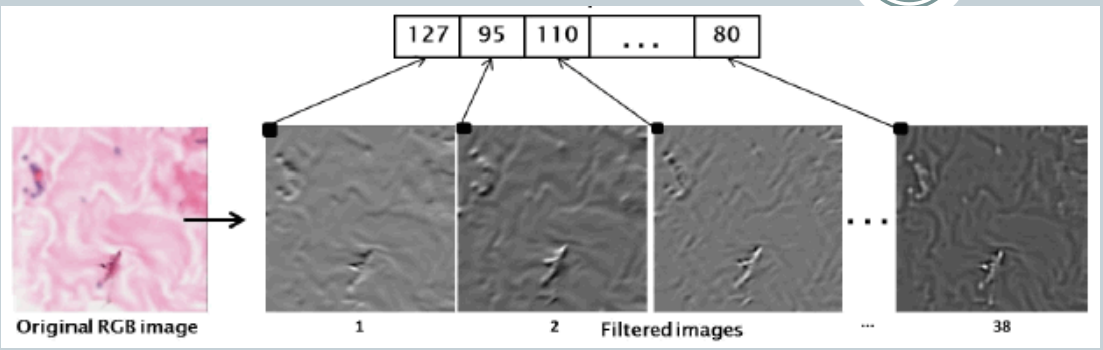
- The filter bank selected to calculate the frequential textons is the maximum response filter bank MR8 which is composed by 38 filters: a Gaussian and a Lapalcian filter and 18 edge and bar filters with 3 basic scales.



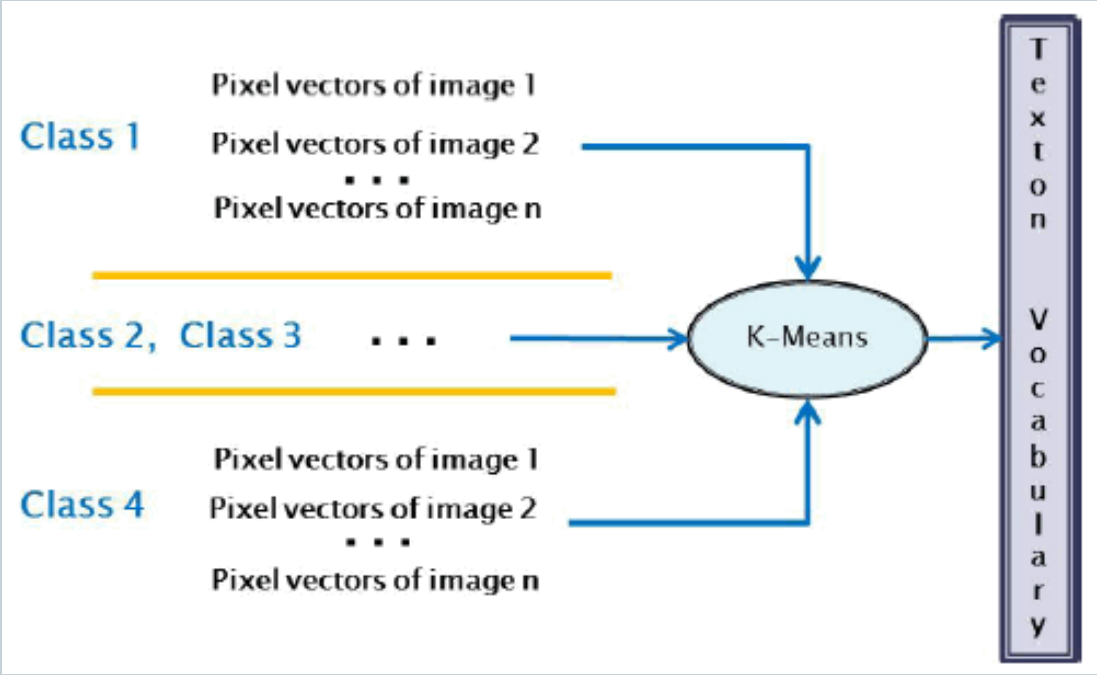
Frequential Textons

- The algorithm used to extract the frequential textons proceeds as follows:
  - 1) The MR8 filter bank is applied over the tissue images so 38 response filters are extracted. Each pixel belonging to the original image is now represented by a 38 dimensional vector.
  - 2) A k-means clustering algorithm is applied over all the pixel vectors. The algorithm creates as many groups as we indicated. ➔ Each new group is characterized by a representative vector called texton ➔ texton vocabulary.

Frequential Textons



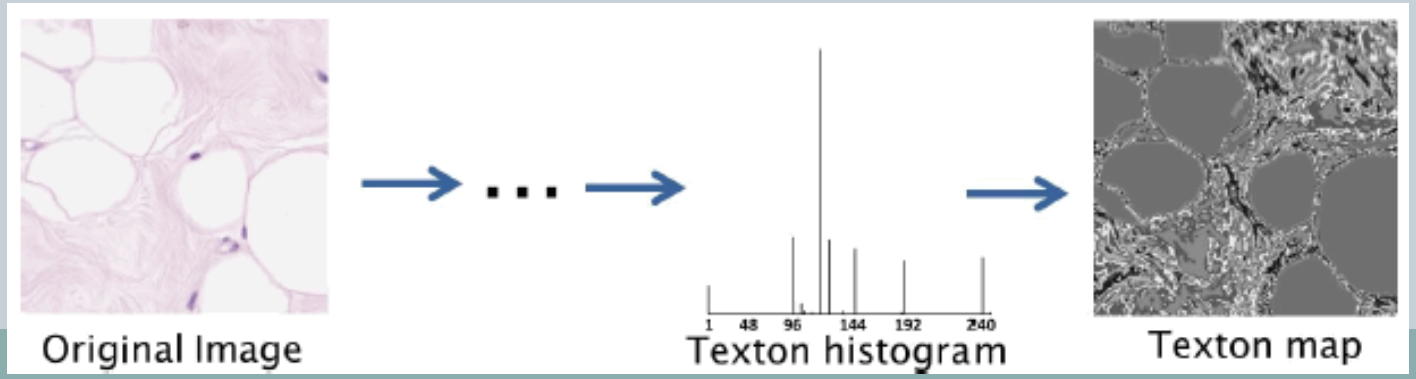
Extracting the 38 dimensional vector of pixel 1 from a RGB image. The pixel vector consists of the 38 first pixel of each filtered image.



Thus, if there are 4 classes and we selected 60 textons for each class ➔
a total of 240 textons are extracted which form our texton vocabulary.

Frequential Textons

- A texton map is generated by each tissue image and the texton vocabulary.
- Maps are generated using the k-nearest neighbours algorithm (kNN) which uses all textons as training samples and predicts the response using each image pixel vector. Finally, the texton selected is the closest texton to the given pixel vector.
- The texton map is a representation of the original image which assigned to each pixel the corresponding texton indices.



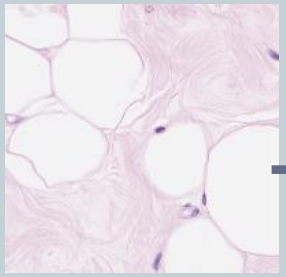
Spatial Textons

- The most important idea is that textons are not filtered by a filter bank. In this case each pixel is represented by the intensity values of an NxN square neighbourhood instead of frequential filters ➔ Thus, the original image is now represented by a N^2 dimensional vector. We use N=3.
- As in the previous type of texton, a k-means clustering algorithm is applied over all the pixel vectors (60 textons per class were extracted).
- Texton maps are computed as the same way as the frequential textons.

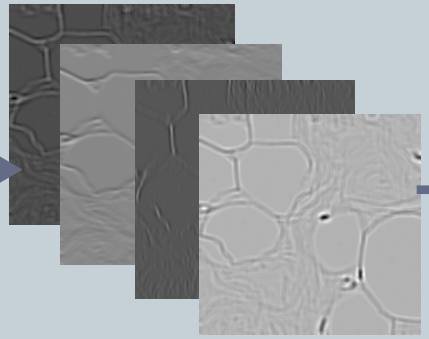
Feature extraction

- Features were extracted calculating the 1st and 2nd order statistics (or Haralick coefficients) on the texton maps for frequential and spatial texton.
- The 1st order statistical descriptors are based on the image histogram. ➔ The histograms can extract statistical values of the gray level image distribution like the mean, the variance or the standard deviation.
- The 2nd order statistical descriptors consider the relationship of the image pixels. They are based on the Grey Level Co-occurrence Matrix (GLCM) of the image. ➔ GLCM are 2nd order histograms that represent the spatial dependence of the image pixels.

Feature extraction



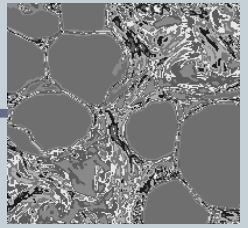
Original Image



N responses

Image pixel vectors (N dimensional vectors)

Texton vocabulary



Texton map

1st and 2nd order statistical descriptors

Feature set

Dimensionality reduction

- The combination of colour models can increase up to 8 times the size of the feature set. ➔ we are dealing with **1928 features**. A dimensionality reduction of the feature vector is needed.

- A forward sequential search method (FSS) was selected:
  - dimensionality reduction of the features
  - removing redundant features
  - Removing irrelevant features

- Reduction of 72.56% of the initial features

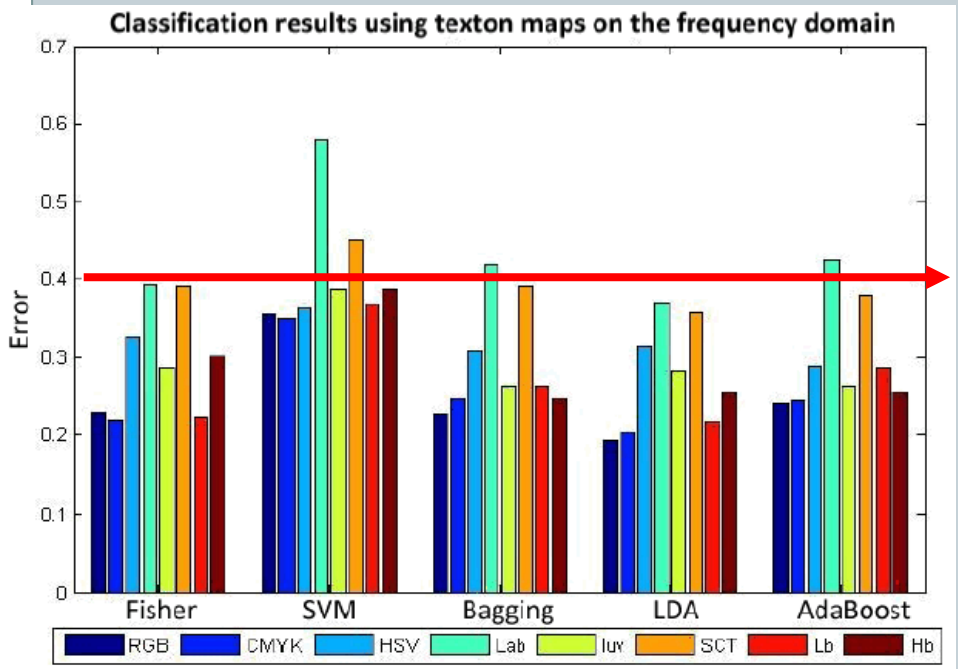
Classification

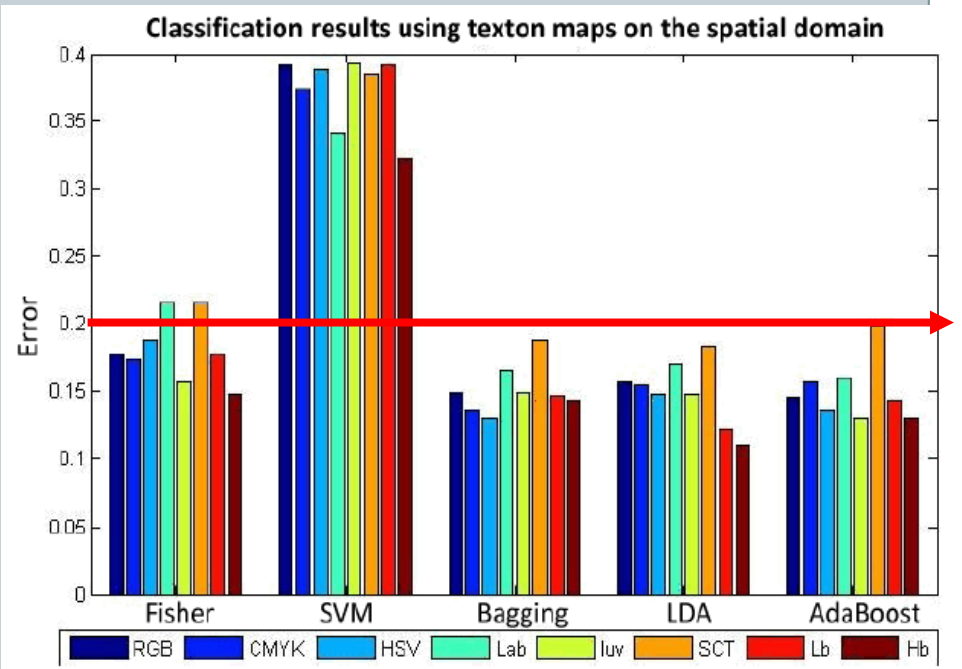
- Classification was performed with each colour model individually, and with the combination of all colour models.
- Several classifiers and Tissue classification was not only performed with we select the AdaBoost classifier.
- AdaBoost assigns weights over the training set. Initially, all the weights are equal but at each iteration the weights of misclassified examples are increased. Thus, in subsequent iterations the weak classifiers will be more focused on these examples.
- Classification performance was tested by 10-fold cross validation and leave-one-out for the best classifier.

Results

Frequency

Spatial





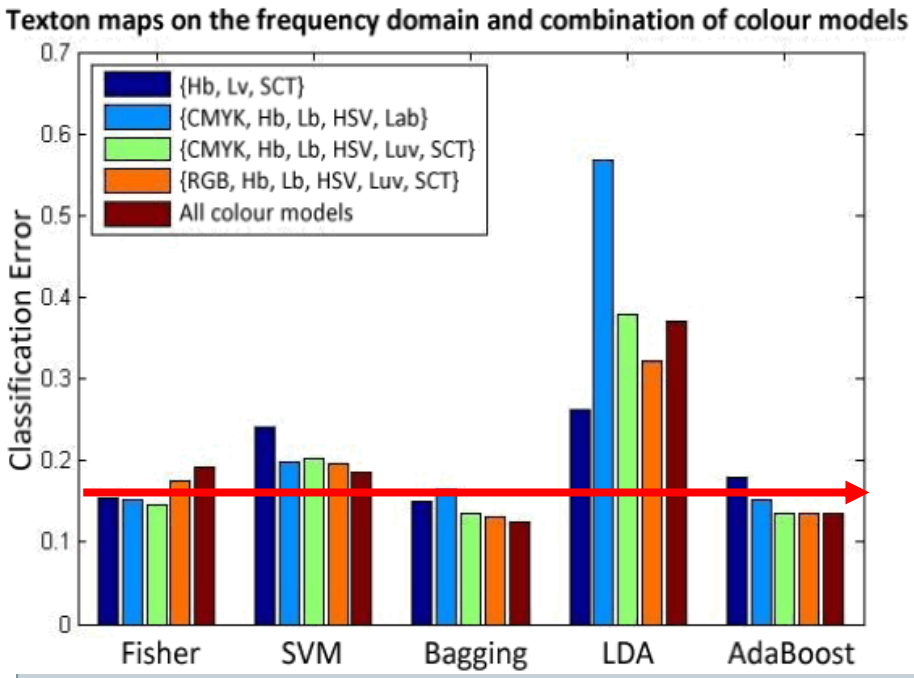
0.4 average error

0.2 average error

Individual Colour Models

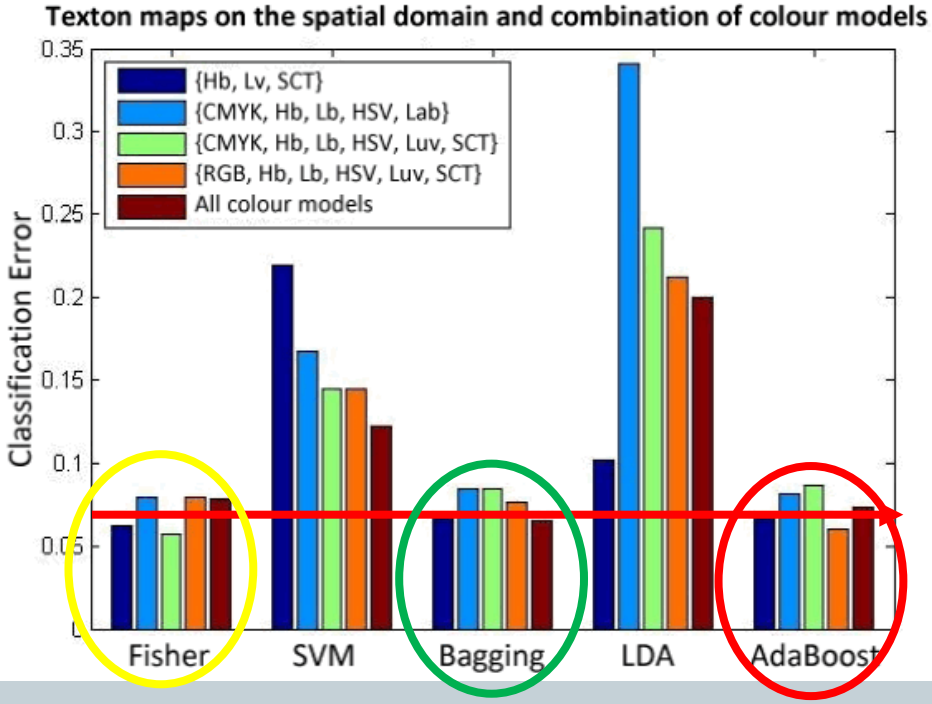
Results

Frequency



0.17 average error

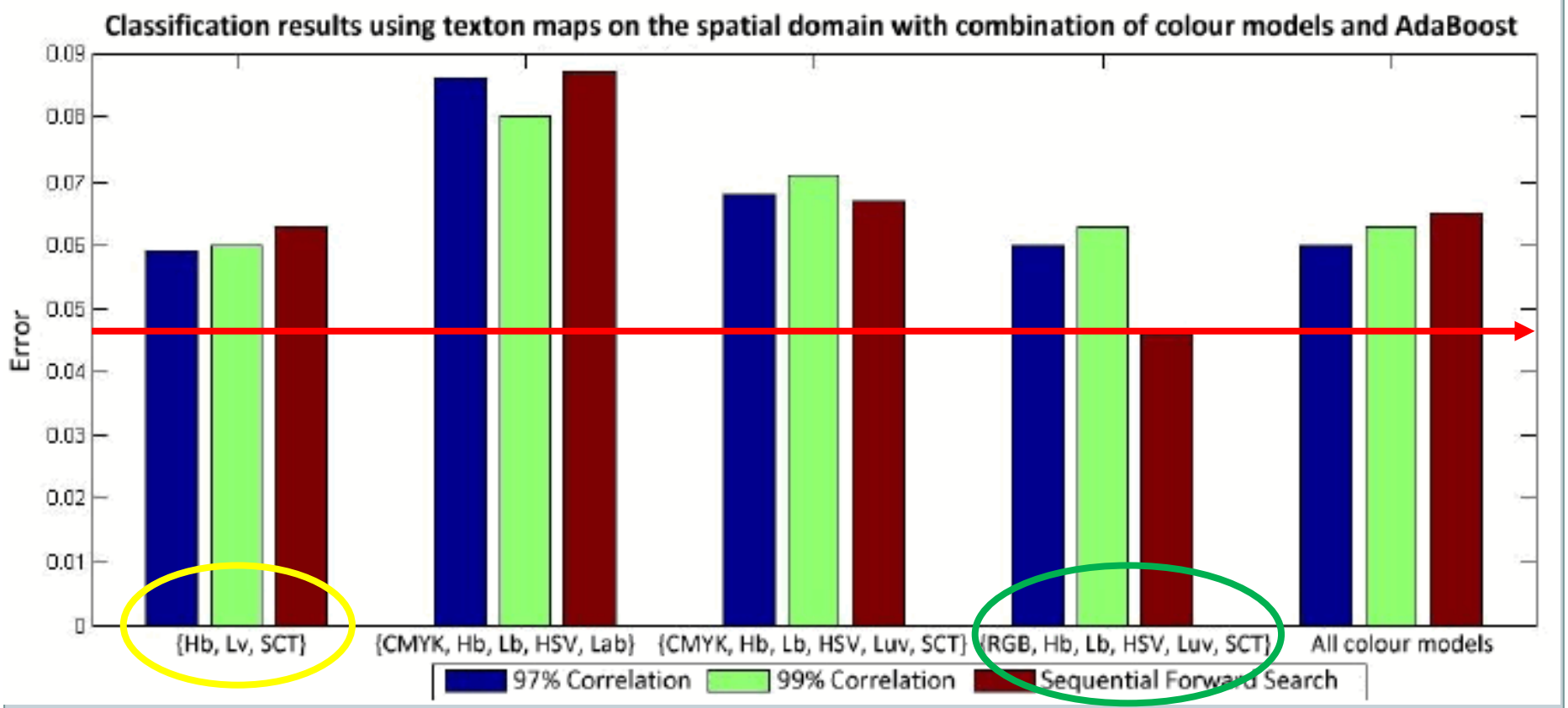
Spatial



0.06 average error

Combination of Colour Models

Results



0.04 average error ➔ 98% accuracy and 96% precision

Spatial Domain Combination of Colour Models and Feature Reduction

Conclusions

- This paper describes a complete study on breast TMA classification based on texton descriptors. A dataset of 628 TMA images divided into four classes was used.
- A suitable combination of colour models and features led us to achieve 98% accuracy and 96% precision using spatial textons, making this study truly valuable in breast TMA classification.
- Although the number of features was large, the AdaBoost classifier takes approximately 79 seconds to perform the training and the test in the classification.
- This procedure is being integrated in a standalone application.

Thank you for your Attention

Questions