IMAGE BASED CHARACTERIZATION OF CIRCULATING TUMOR CELLS

Katrine Brandt Albrektsen, Master Thesis Student
Technical University of Denmark
CytoTrack Aps

MOTIVATION

Images of *circulating tumor cells* (CTCs) can be obtained using CytoTrack. The images have to be manually scored by a trained operator in order to eliminate the false positive images. The scoring can take up to several hours, and it is time consuming and tedious work. The scoring is a qualitative process and both inter- and intra- operator variability must be considered. Some of these problems can possibly be eliminated by using an automatic scoring algorithm.

INTRODUCTION

WHAT IS CTCS?

- Tumor cells from solid tumors found within the blood stream are known as *circulating tumor cells* (CTCs)
- CTCs are very rare, with as few as one CTC among billions of other cells
- The presence of CTCs is linked to poor progression free and overall survival

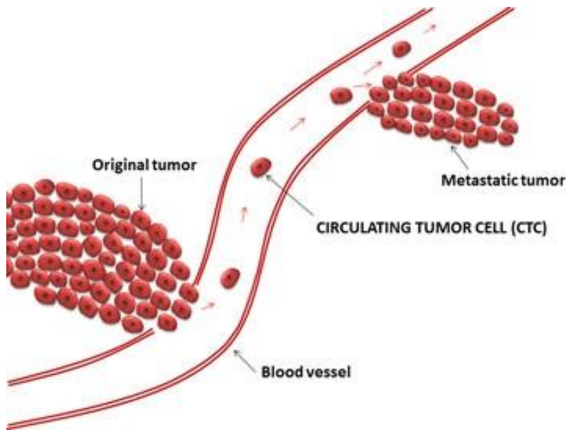

Figure 1: Illustration of how CTCs enter the blood stream and move to foreign sites where a secondary tumor is formed [*T. Hillig et al. In vitro validation of an ultra-sensitive scanning fluorescence microscope for analysis of circulating tumor cells. APMIS, pages 16, August 2013*].

WHAT CAN CTC DETECTION BE USED FOR?

- The number of CTCs found within a blood sample can be used as a prognostic factor or for monitoring the treatment
- Different fluorescent markers can be used to characterize the individual CTCs and help personalize treatment

HOW ARE CTCS DETECTED?

- CTCs can be detected using CytoTrack, which is based on fluorescent microscopy.
- The cells are stained using fluorescent markers, and scanned with CytoTrack
- After scanning the operator will be provided with a catalog of *hot spots*, i.e. areas where there might be a CTC

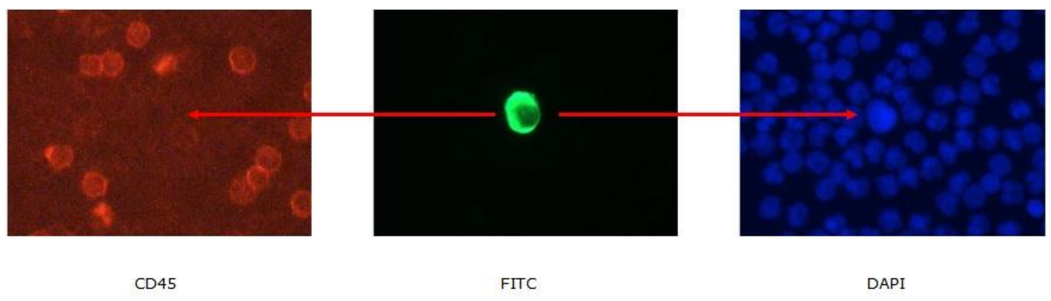

Figure 2: Example of the images obtained with CytoTrack. In this figure a CTC positive image is shown

DEFINITION OF A CTC

For a cell to be classified as a CTC it have to be FITC positive (express cytokeratin), DAPI positive (containing a nucleus) and CD45 negative. Furthermore the cell should have a diameter > 4 µm, and it should have a cell-like morphology

MATERIALS AND METHODS

MATERIALS

- *Image sources:* Patient samples and cell line samples
- *Patients:* Breast cancer patients
- *Cell lines:* originates from breast cancer patients.

METHODS

- *Preprocessing:* In the preprocessing the images are converted to gray scale images and segmented using a fixed threshold
- *Preselection:* Obvious CTC negative images are classified as negative. This is done in order to reduce the amount of data before classification
- *Classification:* Two different methods are tested: **Random Forest** and **Support Vector Machines (SVM)**.
- *Validation:* The performance of SVM and random forest is tested using 10-fold cross-validation. The algorithm to give the highest sensitivity and specificity (without overfitting) is used in the final scoring algorithm.

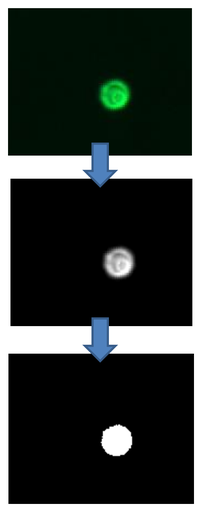

Figure 3: Illustration of the preprocessing on a FITC image

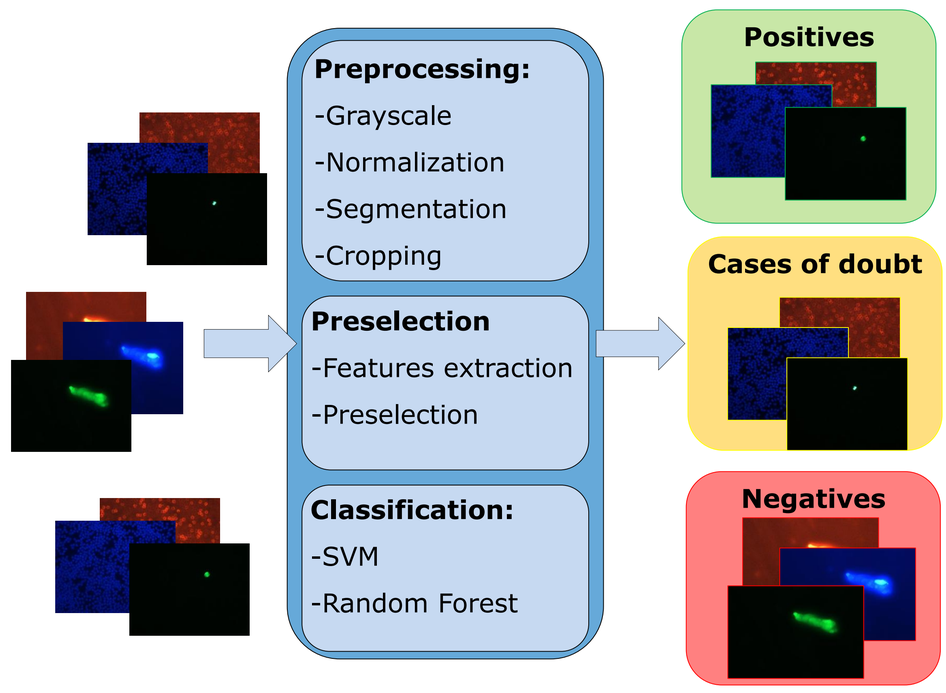

Figure 4: Illustration of the scoring algorithm

RESULTS

The results obtained so far are very promising. To the right are shown the confusion matrix for SVM and for random forest using the parameters that gave the best results in the cross validation (highest sensitivity and specificity without overfitting) for each of the algorithms, respectively. The results are based on cell line samples and the test set used have not been included in the training set.For the moment the images are only classified as either positive or negative (i.e. cases of doubts have not yet been considered).

From the confusion matrices it is possible to compute the *sensitivity* and *specificity* for each of the algorithms:
Random forest: *sensitivity = 0.9512, specificity = 0.9655*
SVM: *sensitivity = 0.9675, specificity = 0.9655*

Results from classification on a seperate testset

RANDOM FOREST		PREDICTED	
		NEGATIVE	POSITIVE
ACTUAL	NEGATIVE	28	1
	POSITIVE	6	117

SVM		PREDICTED	
		NEGATIVE	POSITIVE
ACTUAL	NEGATIVE	28	1
	POSITIVE	4	119

Table 1 & 2: Confusion matrices from two different classification methods.

CONCLUSION

- With the methods presented above both high sensitivities and high specificities are achieved, which make the use of automatic scoring plausible
- The results from the automatic scoring are comparable to the manual scoring
- The automatic scoring is consistent and it is thus possible to eliminate inter- and intra- operator variability