

# Automatic Feature Extraction and Classification of Cell Images for Cytopathology

*A thesis submitted  
in partial fulfillment for the degree of*

Doctor of Philosophy

*by*

GOPAKUMAR G



Department of Earth and Space Sciences  
INDIAN INSTITUTE OF SPACE SCIENCE AND  
TECHNOLOGY  
Thiruvananthapuram - 695547

December 2017

# ABSTRACT

Cytopathology is the analysis at cellular level for disease diagnosis. Every cell has standard morphology and typical count in unit volume constituting the cell signature. Depending on the pathological state of the individual, the signature may change and is the subject of cytopathology. Manual microscopic examination is the gold standard for cytopathology but is a tedious, skill demanding job and suffers from low throughput. Automated microscopy and more recently imaging flow cytometry (IFC) emerged to overcome these difficulties and to standardise the result. However these systems used extensive robotic handling and/or expensive fluid handling mechanisms, making them bulky, expensive and not suitable for resource limited clinics. In our research, we strive for developing very cost-effective point-of-care diagnostics platforms by using off-the-shelf, low-cost components. However the low-cost instrumentation has introduced great challenges in processing the acquired data such as dealing with the focus shift, unlabeled, unstained data and imaging artefacts. We have overcome these challenges by designing, developing and employing sophisticated image analysis and advanced machine learning algorithms. We have proposed processing frameworks for both microscopy and IFC: a framework to automate malaria diagnosis in microscopy and a general framework for processing and classification of cells in IFC. The frameworks include feasible preprocessing, novel cell segmentations, feature extraction as well as classification. We have explored both the possibility of using conventional classifiers (like support vector machine and nearest neighbour) and trending deep learning based classifiers (based on restricted Boltzmann machine and convolutional neural network) and proposed classification techniques even when the availability of labeled data for training is limited. The feasibility of the IFC framework is established by classifying leukaemia cell-lines (K562, MOLT, and HL60).