the diagnostic pathology journal DIAGNOSTIC PATHOLOGY

13th European Congress on Digital Pathology Proceedings, diagnostic pathology 2016, 8:135
ISSN 2364-4893
DOI: http://dx.doi.org/10.17629/www.diagnosticpathology.eu-2016-8:135

Proceedings

PS02.07 | ePoster Session II

Application Of Ki-67 Analysis In A Distributed Computing Infrastructure

M. Strutz*[1], B. Lindequist[1], M. Witt[1], H. Heßling[1], P. Hufnagl[1, 2], D. Krefting[1]

[1]University of Applied Sciences, HTW, Berlin, Germany, [2]Charité - Universitätsmedizin Berlin, Institute of Pathology, Berlin, Germany

Introduction/ Background

Over the last few years, the protein Ki-67 [1] has been established as one of the most important biomarkers for cell proliferation in breast cancer. High Ki-67 values indicate high tumor growth and have direct impact on the patient's treatment. Several automated image analysis methods for identifying Ki-67-positive and negative tumor cells have been presented.

Aims

For small regions of a virtual slide, the Ki-67 analysis can be realized within an acceptable period of time. However, to analyze an entire whole slide image (WSI [2]) most of the current methods are not sufficient yet. On a typical office computer, the processing time of 3,752 tiles, which were extracted from a H–DAB stained WSI, exceeded 24 hours. Therefore, we propose an approach to significantly speed up the process of analyzing entire WSIs by using a distributed computing infrastructure.

Methods

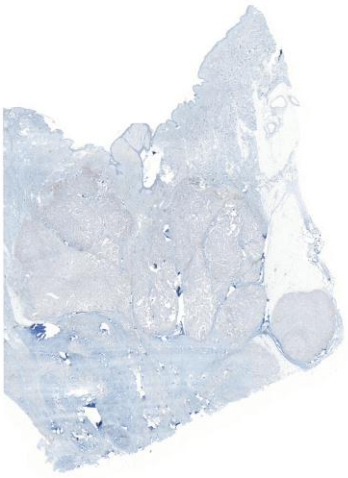
To evaluate the approach, an unmodified and validated [3] [4] analysis software for Ki-67 was deployed on a six node setup supporting two different software engines: Hadoop Streaming [5] and Apache Spark [6]. Both tools support the MapReduce methodology whereas Apache Spark offers alternative programming models. In addition, heat maps visualizing the Ki-67 scores for an entire slide were generated which can provide additional information for clinical research.

Results

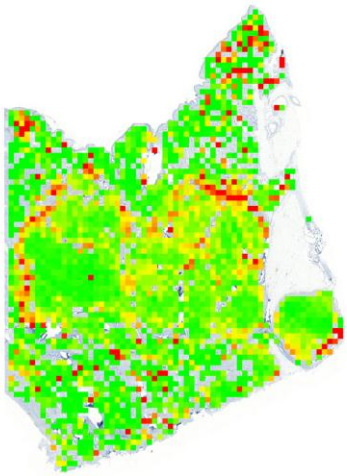
First results on automated and reproducible tests have been produced. By processing 3,752 tiles the speedup turned out to increase linearly with the number of tiles. The overall processing time was improved by a factor of 10, more precisely from 28 hours on a typical office computer to three hours on a distributed environment. Further optimization strategies besides WSI partitioning will be considered. To achieve additional improvements in processing speed, the underlying algorithm of a Ki-67 analysis can be examined with focus on how to adapt it towards distributed processing workflows.

the diagnostic pathology journal DIAGNOSTIC PATHOLOGY

13th European Congress on Digital Pathology Proceedings, diagnostic pathology 2016, 8:135
ISSN 2364-4893
DOI: http://dx.doi.org/10.17629/www.diagnosticpathology.eu-2016-8:135



(a) Overview of a WSI with a dimension of 67,584 x 93,952 pixels



(b) Color-encoded Ki-67 score for all tiles. Each tile has a dimension of 1,024 x 1,024 pixels. If a tile contains no cell nuclei or the nuclei count is below a predefined threshold (threshold equals 50 in this example image) the tile is not encoded:
- green : < 0.15,
- yellow : ≥ 0.15 and < 0.35,
- red : ≥ 0.35

References:

[1] Booth D.G., Takagi M., Sanchez-Pulido L., Petfalski E., Vargiu G., Same jima K., Imamoto N., Ponting C.P., Tollervey D., Earnshaw W.C., Vagnarelli P., Ki-67 is a PP1-interacting protein that organises the mitotic chromosome periphery, eLife 2014, ;3:e01641.

[2] Ghaznavi F., Evans A., Madabhushi A., Feldman M., Digital imaging in pathology: whole-slide imaging and beyond, Annual Review of Pathology: Mechanisms of Disease 2013, 8: 331-359.

3] Klauschen F., Standardized Ki67 Diagnostics Using Automated Scoring – Clinical Validation in the GeparTrio Breast Cancer Study, Clinical Cancer Research 2015, 21(16):3651-7.

[4] Wienert S., Heim D., Saeger K., Stenzinger A., Beil M., Hufnagl P., Dietel M., Denkert C., Klauschen F., Detection and segmentation of cell nuclei in virtual microscopy images: a minimum-model approach, Scientific Reports 2012, 3:503.

[5] Hadoop Streaming allows to create and run Map/Reduce jobs with any executable or script as the mapper and/or the reducer [Internet], Apache Software Foundation (US), 2016. Available from: https://hadoop.apache.org/docs/current/%20hadoop-streaming/HadoopStreaming.html

[6] Zaharia M., Chowdhury M., Das T., Dave A., Ma J., McCauley M., Franklin M.J., Shenker S., Stoica I., Resilient Distributed Datasets: A Fault-Tolerant Abstraction for In-Memory Cluster Computing [Internet], 2012, Available from: https://people.csail.mit.edu/%20matei/papers/2012/nsdi_spark.pdf~