Prostate Cancer Analysis Using MapReduce and Unsupervised Learning Methods

Zoya Yasin, Master of Science in Computer Science

University of Dublin, Trinity College, 2022

Supervisor: Professor Khurshid Ahmad

In recent times there have been many advancements in the field of digital pathology, owing to the use of Whole Slide Imaging (WSI) technology and the availability of high-resolution imaging sensors. However, there is still a lot that needs to be done to revolutionize diagnosis. There has been a growing demand for faster and more affordable diagnoses of chronic diseases like cancer, and the field of medicine could benefit from the use of automated systems that aid the screening and evaluation of tissue samples taken from patients. Additionally, the general trends in pathological analysis favour shifting to a digital platform, but the overwhelming size of datasets created by the generation of high-resolution medical images is a huge deterrent. In accordance with said requirements, this study has proposed an approach to medical image analysis with the help of a parallel computing framework (MapReduce) and unsupervised clustering methods. The experiments in the proposed system utilize a dataset of prostate tissue images. Due to parallel processing abilities, this system can process and extract quantifiable features from the image in significantly smaller amounts of time than traditional systems. The features extracted using the above framework are robust to changes in positioning and noise and eliminate redundancy of information presented by them. The system is successfully able to distinguish between two separate categories of prostate tissue based on learning from the extracted feature vector.