AbstractID: 8601 Title: Grid-Computing for Optimization of CAD

Purpose: Pilot study to utilize large scale parallel grid computing to harness the nationwide cluster infrastructure for optimization of medical image analysis parameters.
Methods: A previously developed CAD scheme for mass lesions in mammography was ported onto the grid computing environment by wrapping the algorithm code with the virtual data language (VDL). The CAD scheme was then configured into a parallelizable workflow by the grid-software. The workflows were executed using two test clusters (in Santa Monica, CA and Chicago, IL) consisting of over 220 dual-CPU nodes combined. Using the grid-environment workflow, parameter sweeps were conducted for lesion segmentation settings based on radial-gradient-index (RGI) methods. Specifically, the Gaussian width (GW) used in initially filtering lesion images for segmentation was varied by increments of 1 mm from 1 to 60 mm. For each GW sweep the entire 850 biopsy-proven mass lesion database (411 benign, 439 malignant) was analyzed. In each, 29 different mathematical descriptor features were calculated, followed by feature selection and merging with linear discriminate analysis. Diagnostic performance was estimated by ROC analysis by calculating AUC (from PROPROC) values based on both individual features alone, and merged. For merged classifiers, AUC values were found using round-robin case-by-case removal and replacement.
Results: Computation jobs requiring over 30 CPU hours on a single lab computer were completed in approximately 35 minutes in this preliminary study. Merged AUC values increased from 0.50 (std.err.=0.018) at GW of 1mm with, to 0.81 (std.err.=0.015) at 10mm GW, with relative plateaus across the rest of the parameter space to 60mm.
Conclusion: The parameter space sweep in GW identified trends in individual feature performance as well as merged results. Large scale, computationally intensive image analysis can be carried out in a timely fashion, feasible for expedited experimental discovery, as well as for more thorough future statistical analysis.