Paper Title: High Performance Computing for Detecting Complex Diseases using Deep Learning

Paper Link: https://ieeexplore.ieee.org/document/9194158

1 Summary

1.1 Motivation

The paper is motivated by the limitations of single linear SNP (single nucleotide polymorphism) effect studies in the analysis of genotype-phenotype relationships for a variety of diseases in Genome-Wide Association Studies (GWAS). The study highlights the necessity of investigating higher-order interactions in order to gain a deeper understanding of the genetic relationships with various diseases, specifically SNP pairwise interactions and epistasis.

1.2 Contribution

In this paper, Parallel Deep Learning (PDL) with Map/Reduce is introduced and tested on real and simulated Rheumatoid Arthritis data for multi-node processing on a supercomputer. Through metrics like accuracy, TPR, Spc, and FDR, it compares PDL's performance and shows that it is robust across disease models.

1.3 Methodology

Numerous SNP-SNP interaction algorithms are examined in the paper. The methodology consists of two phases: Phase 1 uses parallel Fork/Join to preprocess SNPs, Phase 2 uses Parallel Deep Learning on a supercomputer with H2O. The datasets comprise an actual Rheumatoid Arthritis dataset as well as simulated data with twelve 2-locus disease models.

1.4 Conclusion

The study concludes that even in the presence of genetic heritability variations, PDL exhibits robust performance in simulated data across various disease models. Using a real RA dataset, PDL is applied to show important interactions and support biological gene analysis.

2 Limitations

2.1 First Limitation

The paper recognizes that model-based challenges impact PDL's average power, suggesting a possible constraint in some situations.

2.2 second Limitation

Linkage disequilibrium and missing genotypic data are two examples of factors that could impact some of the algorithms discussed in the paper.

3 Synthesis

The study demonstrates Parallel Deep Learning's potential for comprehending SNP-SNP interactions by integrating various algorithms, introducing it, and evaluating it on both simulated and real datasets. The thorough assessment contrasts it with current approaches, and the conclusion lists accomplishments and suggests areas for further investigation.