Proposal

**Author:** Gregory Schwing, Wayne State University School of Medicine, Detroit, MI

**Mentor:** Chindo Hicks, Department of Genetics, School of Medicine, Louisiana State University Health Sciences Center – New Orleans, Louisiana

**Date:** 10/24/18

**I. Title** : A Novel Bioinformatics Approach to Estimate Hazard in Pediatric Glioma

**II. Specific Aims**

In conducting this study, we will accomplish the following specific aims:

Specific Aim 1. Test the ability of neural networks, specifically the “cox-nnet” program, to estimate Hazard score/ratio in pediatric glioma. The results will be compared against other more established prediction algorithms such as Cox proportional hazards (Cox-PH) methods with LASSO or ridge penalization.

Hypothesis 1. Neural networks better model biological signal and better predict hazard than other approaches due to the similarity they show to biological signal processing.

Specific Aim 2. Search for novel predictors of hazard in pediatric glioma, specifically surrogate variables generated by the model.

Hypothesis 2. There exist surrogate variables which better predict time to event than individual gene expression. These variables represent interactions between genes which reduces dimensionality and increases survival sensitivity.

**III. Background**

Pediatric high-grade glioma is a major public health problem. Over the last 18 years considerable effort has been devoted to discovery of molecular prognostic markers and targets for the development of novel therapeutics: Recently there has been a a growing interest in the use of neural networks. [1] **Cox-nnet** is a program developed by the Garmire Group at the University of Michigan. [2] The software is a Artificial Neural Network with two or more layers that performs Cox (Proportional Hazards) Regression in the final layer. The neural network learns via a process of training, optimizing, and testing to find hidden connections between variables to best estimate Hazard score/ratio. These hidden connections form surrogate variables which may complement pathway analysis as a bridge between genotype and phenotype.

**IV. Research Methods**

Study Design: Metaanalysis of gene expression from Affymetrix arrays, Agilent arrays, and RNASeq platforms with Overall Survival as the primary endpoint and Progression Free Survival as the secondary endpoint.

Study Subjects: The study subjects are as follows – 102 subjects with Affymetrix U133 Plus 2.0 data, 67 subjects with Agilent Whole Human Genome 44k data, and 653 subjects with RNASeq data.

Analyses: The analyses to be conducted fall into two classes – model performance and biological insight. The model performance analyses are as follows 1) C-IPCW (inverse probability of censoring weighted), 2) Harrell’s concordance index, 3) log-ranked p-value, and 4) Integrated Brier Score. These scores will be used as metrics to compare the performance of the neural network against the other algorithms. The biological insight analyses are as follows – 1) Gene Set Enrichment Analysis, 2) 3) Causal Variant Analysis, and Cox (Proportional Hazard) Regression.

**V. Timeline of Research Project**

**Month**

**Activity** 1 2 3 4

Data Access Compliance X

and Data Retrieval

Preprocessing

& X

Cox-nnet Run

Other Algorithms Run X

Post-experimental

Analysis and Poster X

Presentation at AANS

**VI. Literature Cited**

**1.** Vohradsky, Jiri. "Neural network model of gene expression." *the FASEB journal* 15.3 (2001): 846-854.

**2.** Ching, Travers, Xun Zhu, and Lana X. Garmire. "Cox-nnet: An artificial neural network method for prognosis prediction of high-throughput omics data." *PLoS computational biology* 14.4 (2018): e1006076.